On the origin of ‘bloopergenes’: unraveling the evolution of the balanced lethal system in *Triturus* newts

Manon de Visser

Colofon

ISBN: 978-94-6473-725-7

De Visser, 2025. On the origin of 'bloopergenes': unraveling the evolution of the balanced lethal system in *Triturus* newts.

PhD thesis, Leiden University, The Netherlands

This research was carried out at the Institute of Biology Leiden (IBL) at Leiden University, Leiden, The Netherlands, in collaboration with Naturalis Biodiversity Center (NBC), Leiden, The Netherlands. The project was financially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 802759).

Cover design: Design by Manon de Visser, using images by; Bas Blankevoort (embryo drawings, adjusted) and Erik-Jan Bosch (male *Triturus dobrogicus* drawing, adjusted) – both under the Open Content License of Naturalis Biodiversity Center (© CC BY-NC-ND 4.0), Pixabay.com (grey clouds, adjusted) under the Pixabay Content License, and Manon de Visser (back cover photo of *Triturus* eggs folded in aquatic vegetation).

Lay-out: Typeset by Manon de Visser, using Microsoft Word Version 2411.

Printing: Ipskamp Printing, Enschede, The Netherlands.

Photos: The page-filling photos per each chapter/appendix were used with the written permission of the photographers.

On the origin of ‘bloopergenes’: unraveling the evolution of the balanced lethal system in *Triturus* newts

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op woensdag 5 maart 2025

klokke 13:00 uur

door

Manon Chantal de Visser

geboren te Spijkenisse, Nederland

in 1993

Promotor:	Prof.dr. M.K. Richardson
Co-promotor:	Dr. B. Wielstra
Promotiecommissie:	Prof.dr. A.H. Meijer
	Prof.dr. H. de Winde
	Prof.dr. M. Schilthuizen
	Dr. M. Bosse *Vrije Universiteit Amsterdam en Wageningen University & Research*
	Dr. K. Peijnenburg *Universiteit van Amsterdam en Naturalis Biodiversity Center*

"하시죠. 영화에서 보신 적이 있을 겁니다 – 러시아 룰레."

"Shall we start? I'm sure you've seen this in the movies – Russian Roulette."

– Gong Yoo as The Recruiter in *Squid Game* Season 2

12-day-old Triturus *embryo (X-radia microCT scan)*

Table of contents

Triturus *newts in a glass cuvette*

© Christos Kazilas
(permission to use picture granted)

Chapter 1 - General introduction

Manon de Visser[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands
2. Understanding Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands

Evolution by natural selection is the theory in biology that explains the origin and diversity of all life on earth [6, 7]. The process of evolution, in which natural selection favors traits that enhance the survival and reproductive output of individuals in populations, is inherently not goal-oriented, nor predetermined [8, 9]. Despite this, my observation is that adaptations are often categorized as evolutionary ‘successes' or evolutionary 'failures’, depending on their relative fitness consequences [10, 11]. Strictly speaking, this is a distinction which I think translates the randomness of the evolutionary process into understandable, human-defined constructs (since nothing in evolution is truly an achievement or a blunder: there is no end goal). Nevertheless, I will use such definitions throughout this dissertation, as it would be very hard to discuss evolutionary mechanisms otherwise.

Evolution is thoughtless, lazy and imperfect

While evolutionary successes are broadly recognized and comprehended, evolutionary failures remain misunderstood, understudied, and likely overlooked in nature [6, 10, 12, 13]. Natural selection acts on the phenotypic level: it works against individuals that are less fit to their environment, and it works in favor of individuals that are more fit to their environment [6, 7, 11]. In other words: the individuals with so-called favorable phenotypes are the ones that are more likely to survive, reproduce, and pass on their genetic makeup to the next generation (i.e., they have a higher fitness), compared to individuals with unfavorable phenotypes [14, 15].

Today it is known that, in sexually reproducing diploids that generally carry two gene copies, the mechanism of inheritance often leads to a change in frequency of the occurrence of heritable traits over the course of successive generations [16]. This shift is influenced not only by the works of natural selection, but also by the (combined) works of genetic drift, recombination, gene flow, and mutation – all of which influence the genetic variability necessary for adaptive potential [15, 17]. Overall, the favorable mutations, or more specifically the favorable ‘alleles’ (gene variants), are positively selected for, whereas deleterious (including lethal) alleles are selected against. This is referred to as directional and purifying selection, respectively [11, 14, 18].

The effect of purifying selection in particular is most strong when an unfavorable allele appears lethal, especially in the pre-reproductive phase as it prevents an individual from reproducing altogether. Therefore, dominant lethal alleles (of which only one needs to be present in a diploid organism to be expressed phenotypically) are hardly detected in nature because of their rapid elimination from the gene pool (Fig. 1A). Recessive lethal alleles, on the other hand, can only be negatively selected against if their frequency in the population is high enough that they start to appear in a homozygous state [i.e., they will

only be phenotypically expressed if a diploid organism carries two similar copies of them; 11, 14, 15]. Thus, these are the types of alleles that are easily purged in populations that contain a relatively high frequency of them, but that can really only be eliminated to the point that these mutations must linger on within the population at very low frequencies [assuming an infinite population size, or more particularly, assuming the absence of an effect of genetic drift; 15, 19, 20]. This is because they will remain concealed by the heterozygous masking of the carriers, meaning that - theoretically - the frequency of recessive, harmful alleles is subdued by natural selection, but can never reach zero (Fig. 1B). One would think that, under such natural laws that automatically lead to the suppression of disadvantageous variants and to the enhancement of beneficial ones, nature will always end up 'having the upper hand' in the long term. In other words: nature would not allow genetic anomalies or genetic diseases (i.e., evolutionary failures) to prevail in a population, at least not on abundantly. However, by means of this dissertation, I intend to show you that this is not necessarily true.

To get straight to the point: the most remarkable example of an evolutionary failure is an extremely deadly one. Imagine an extraordinary situation in which a diploid organism possesses two different versions of a chromosome with unique, recessive lethal alleles that are compensated for by the functioning gene copy on the other chromosome version (Fig. 1C). In such a situation, natural selection will be unable to purge these harmful alleles and will reach what could be considered an 'impasse', as on these terms all homozygous offspring are inviable, whereas all heterozygous offspring depend on both chromosome versions for their viability. Subsequently, these heterozygous individuals grow into adults and reproduce, repeating the cycle. This is called a 'balanced lethal system' [21]: a genetic anomaly which will always lead to the reproductive output being cut in half. While the maintenance of such a system is relatively easily explained in theory, the details on the origin of it are all the more puzzling (I will arrive at the details on this later).

In the grand scheme of things, a balanced lethal system can only be the result of evolution by natural selection being a thoughtless, lazy, and imperfect process. As the famous biologist Richard Dawkins described: *"Natural selection, the blind, unconscious, automatic process which Darwin discovered, and which we now know is the explanation for the existence and apparently purposeful form of all life, has no purpose in mind. It has no vision, no foresight, no sight at all. If it can be said to play the role of watchmaker in nature, it is the blind watchmaker"* [22]. The most famous example of a balanced lethal system has been described in the wild, in a specific group of amphibians, and forms the basis of this research dissertation.

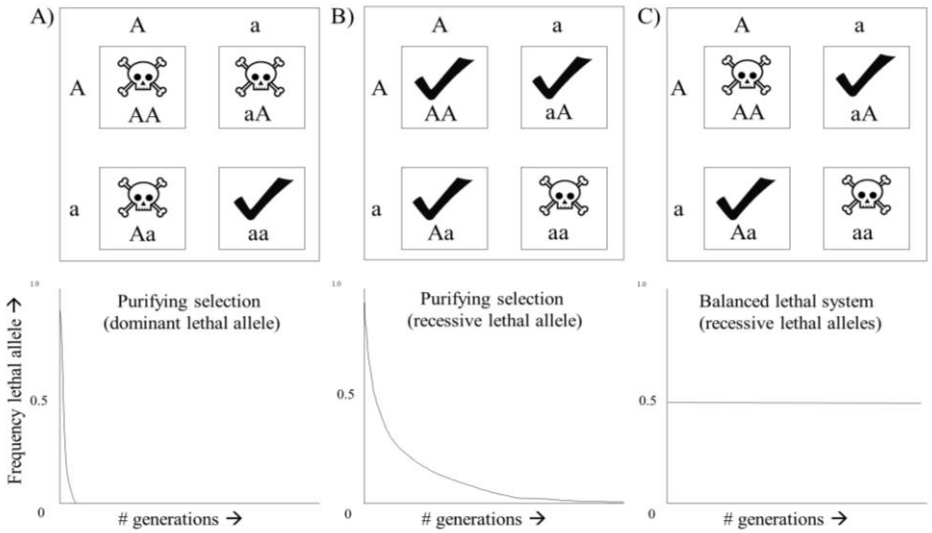


Figure 1: *Schematic representations of the frequency of harmful alleles over a certain number of generations under three different scenarios observed in nature (assuming infinite population sizes). Top: Punnett squares. Bottom: Allele frequency graphs.* ***A)*** *Purifying selection causes dominant, lethal alleles to be quickly eliminated from the gene pool (hence, they are hard to detect).* ***B)*** *Purifying selection causes recessive lethal alleles to be purged as much as possible (however, in theory they could never reach a frequency of zero).* ***C)*** *A balanced lethal system is an exceptional situation in which only heterozygotes are viable and natural selection is prevented from purging the lethal alleles (which will be maintained in the population at a frequency of 0.5).*

Of all vertebrates, amphibians are natural outcasts

Amphibians form a diverse group of vertebrates with traits that have fascinated scientists for centuries. They have clearly been evolutionarily successful in adapting to both aquatic and terrestrial environments, hence their name ("amphibian" comes from the ancient Greek word "*amphibios*," which means "both kinds of life" or "living a double life"). For example, they are generally able to respire through their lungs, as well as through their skin [23, 24]. They also show various defense strategies, an exceptional example being the epimorphic regeneration abilities displayed by some species after losing a limb, tail, or even organ, in order to avoid predation [25-27]. Despite these evolutionary successes, amphibians belong to one of the most vulnerable vertebrate groups when it comes to anthropogenic environmental change and pollution [28-31]. Semi-aquatic salamanders, for example, are particularly susceptible to habitat changes due to their limited dispersal abilities [32, 33]. This vulnerability has led to some surprising adaptive behaviors, such as

newts hibernating in unconventional places such as waterfowl nests [observed in the canals of Leiden, The Netherlands; 34]. However, much remains unknown about the mysterious terrestrial lifestyle of pond-dwelling salamanders, including their land use and hibernation habits [although a variety of methods are tested nowadays to research this, one being canine detection; 35, 36-39].

Another remarkable trait of salamanders that I would like to highlight (before I will explain more about the evolutionary mystery that is the balanced lethal system), is their genomic gigantism. Compared to other animals, salamanders have incredibly large genomes that are full of non-coding repetitive DNA and that, depending on the species, can be over ten times the size of that of a human [40-45]. So far, it seems that natural selection has simply tolerated the genomic expansion observed in the clade of the salamanders (a result of neutral evolution). However some studies do suggest that the bigger the genome of a salamander, the less successful it is at limb regeneration – even while it is simultaneously suggested that larger genomes delay cellular differentiation more effectively, which should actually facilitate the regenerative process [45, 46]. Clearly, this is paradoxical, and scientists have yet to decide whether genomic gigantism in salamanders can be considered an evolutionary success, an evolutionary failure, or perhaps something in between. One thing scientists have agreed on, however, is something that I as a salamander scientist can fully stand by: these gigantic genomes are a hurdle to investigate in-depth (a topic I will revisit later in this dissertation).

Triturus*: survival of the ~~fittest~~ *least doomed

At this juncture, I would like to revisit balanced lethal systems. Among all amphibian peculiarities, the most striking one has to be the balanced lethal system found specifically in the crested and the marbled newts of the genus *Triturus* [21, 47-49]. Known also as 'chromosome 1 syndrome', this condition in *Triturus* newts involves two distinct versions of chromosome 1 - the longest of the 12 pairs that they possess – which has been linked to extreme egg mortality rates [48, 50]. As visualized in Fig. 1C, this genetic syndrome causes half the offspring to succumb in the face of natural selection, regardless of the environmental circumstances [47, 48, 50, 51] – something that could be considered the biggest evolutionary failure of all. However, despite seeming unnatural, balanced lethal systems are observed in some other organisms besides *Triturus* newts as well, such as in certain species of fruit flies [52], flour beetles [53] and flowering plants like the rock isotome [54], sundrop [55], boat lily [56] and the evening primrose known as 'groundsmoke' [57]. In scientific literature, however, the example of *Triturus* newts is the most famous, as it is described most recently and in greatest detail as compared to the other cases known.

The case has puzzled scientists for at least two hundred years. For instance, the zoologist Mauro Rusconi already described this phenomenon in his book '*Amours des Salamandres Aquatiques*' as follows: *"…although this embryonic stage is a pleasure to look at for the researcher, it is dangerous for the little creatures themselves, because almost precisely half of them die …"* [freely translated from French; 58]. Here, Rusconi explains that half of all *Triturus* embryos perish while still inside the egg, and he refers to the specific moment in the embryonic development that is detrimental (more on that later).

Because of their high fitness disadvantage, balanced lethal systems seem to defy evolutionary theory. The selective removal of deleterious variants that occurs through the act of natural selection is usually important to rid natural populations of such detrimental fitness effects [6, 59]. Thus, in the light of evolution it is expected that a balanced lethal system can never become fixed, as individuals that do not carry the balanced lethal system at the moment of origin would simply outperform carriers. Also, any emerging balanced lethal system would, in theory, be broken down before fixation, as sexual reproduction is associated with allele segregation and recombination [11, 14, 60, 61]. Especially the effect of recombination in the homozygotes would lead to a disconnection of lethal alleles on both chromosome forms, causing the balanced lethal system to evanesce instantly – which implies that recombination must be suppressed somehow [21]. How exactly this works, is explained in the remaining Chapters of this dissertation, but the crux of the matter is as follows: the genetic structure of a balanced lethal system renders selection against it impossible, because of the compensating effects of the two chromosome versions [62-64], and this leads to an 'eternal' wasting of 50% of the reproductive output. A huge failure, that does not appear to follow the winning '*survival of the fittest*' principle, but rather a losing '*survival of the least doomed*' fallacy, and it makes one wonder: how can something this wasteful originate in nature?

The (brutally) costly courtship & oviposition of *Triturus* newts

The geographical distribution of *Triturus* species ranges from western Europe to western Asia [see Fig. 2; 1, 2]. Currently, ten *Triturus* species are recognized [see Table 1; 2]; the Anatolian crested newt (*T. anatolicus*), the Italian crested newt [*T. carnifex*, comprising of two diverged populations that might be different species, but have so far not been described as such; 4]), the northern crested newt (*T. cristatus*), the Danube crested newt (*T. dobrogicus*), the Balkan crested newt (*T. ivanbureschi*), the southern crested newt (*T. karelinii*), the Macedonian crested newt (*T. macedonicus*), the marbled newt (*T. marmoratus*), and the pygmy marbled newts [*T. pygmaeus* and the newly described *T. rudolfi*; 5]. The most recent common ancestor of *Triturus* is thought to have lived around ± 24 million years ago [1, 2, 65], suggesting that the balanced lethal system originated also

that long ago, as all known *Triturus* species suffer from chromosome 1 syndrome.
In all species of *Triturus*, reproduction starts at about 2-4 years of age [33], and it is a costly process. Every year during the aquatic phase the males develop a crest, which is highly denticulated in crested newts, while smooth in marbled newts [66]. They use this breeding 'cloak' (see Table 1) to impress the females under water in an elaborate courtship dance [described in detail in; 67]. Once a male and female *Triturus* newt meet in a breeding pond, the male immediately starts dancing: while doing this, he not only grasps her attention, but he also spreads his pheromones her way [66].

The female ultimately makes a choice: if she is not in the mood, she swims away, whereas if she is interested, she calmly waits [67]. This initiates the next stage of the courtship dance: the display. The display includes several, alternating behaviors performed by the male, such as 'rocking and whipping' [shifting between lashing out with his tail and doing a handstand kind of pose; 67], 'leaning in' [hanging over the female, showing off the shapes and colorations of his body; 33, 67] and "cat buckling" [arching his back, much like a cat does when stretching or when threatened; 67]. If the female ultimately touches his tail with her snout, the male releases a spermatophore which the female will then take up with her cloaca to fertilize her eggs internally [66, 67].

Figure 2: *Map showing the general distribution of the species of* Triturus *[adjusted from; 1, 2]. The pins refer to the populations that the initial founders of the breeding lines of which I use samples throughout the studies included in this dissertation, originated from. Note that two genetically distinct populations of* T. carnifex *are included [4], and that* T. rudolfi *was described only in 2024 as a new species [5] and therefore is not shown here, nor included in the studies of this dissertation.*

After the male has done his fair share of hard work, the female takes over: she will carefully lay her fertilized eggs – one by one – in between the leaves of the water vegetation [33, 50]. This oviposition strategy, in which she uses her hind limbs to fold, wrap, and press the

leaves around each egg, presumably serves to spread out and conceal the offspring as to hide them from external forces and predators and thus maximize fitness [and also, no further parental care exists; 67]. A *Triturus* female may lay around 200 eggs to 400 eggs in this manner per breeding season [33, 67]. It is, in a way, brutal that half of those eggs – in which both parents have invested so much resources – are not even viable.

Table 1: Species affected by chromosome 1 syndrome (genus *Triturus*). The pictures are taken by collaborator Mr. M. Fahrbach and shows the adult male morphology during the breeding season for each species [except the recently described *T. rudolfi*; 5].

Phenotype	**Common name**	**Scientific name (*Triturus*)**
	Anatolian crested newt	*T. anatolicus*
	Italian crested newt*	*T. carnifex*
	Northern/Great crested newt	*T. cristatus*
	Danube crested newt	*T. dobrogicus*
	Balkan crested newt	*T. ivanbureschi*
	Southern crested newt	*T. karelinii*
	Macedonian crested newt	*T. macedonicus*
	Marbled newt	*T. marmoratus*
	Pygmy marbled newt	*T. pygmaeus*
Not available in this series	Pygmy marbled newt	*T. rudolfi*

** Note that, in some research Chapters of this dissertation, two genetically distinct* T. carnifex *populations are included [4].*

Embryonic arrest occurs in half of all *Triturus* eggs

The two versions of chromosome 1 that *Triturus* newts possess are known to be morphologically distinct [48, 68], where in one version (called '1A') the long arm is longer than in the other version (dubbed '1B'). No chiasmata are formed between the heteromorphic regions and these do not recombine [68, 69]. Individuals that are heterozygous for the two versions survive, whereas both types of homozygotes die during the late tail-bud stage of embryonic development [51, 69-72]. In existing literature, 1A1B (/1B1A) embryos are classified as 'viable', 1B1B embryos as 'fat-tailed', and 1A1A embryos as 'slim-tailed'. These labels are based on supposed differences in tail-bud stage morphology, where fat-tailed embryos would show anomalies, whereas both slim-tailed and viable embryos would appear normal in shape [47, 51, 72].

Embryonic development is driven by strict processes that regulate pattern formation, morphogenesis, cell differentiation and growth [73]. Before homozygote *Triturus* embryos fall victim to the deadly chromosome 1 syndrome, they seem to successfully complete the embryogenic phases of cleavage, gastrulation and neurulation, although a yolk plug seems to interfere with the caudal fusion of the neural folds in some embryos [69, 71, 72].

But then, when the first organs start to develop (a phase called 'organogenesis'), more problems arise. Organogenesis is divided into what is called the 'tail-bud stage' and the 'larval stage', the latter being the last stage to complete before finalizing organogenesis and proceeding towards hatching [51, 71, 74]. During the early tail-bud stage, development of inviable embryos slows down and arrest occurs during the late tail-bud stage in all species, including interspecific hybrids [48, 51, 70-72].

Embryonic arrest kicks in during the phylotypic stage, an ontogenetic period that shows the highest level of conservation among vertebrates (Fig. 3), and is typically referred to as the thin part of the 'hourglass model' [75]. This is the ontogenetic period during which the basic body plan is laid out [73, 75-77], and it corresponds to the tail-bud stage of *Triturus*. The most apparent embryonic changes that take place during the tail-bud stage are the prolongation of the notochord, the development of somites, and the generation of the neural tube [72-74].

Although amphibian embryos are often used as a model to study developmental biology in general and to learn more about the mechanisms behind limb generation [27, 45, 78-81], nobody so far has investigated the mutations that cause chromosome 1 syndrome, and how these could have possibly evolved.

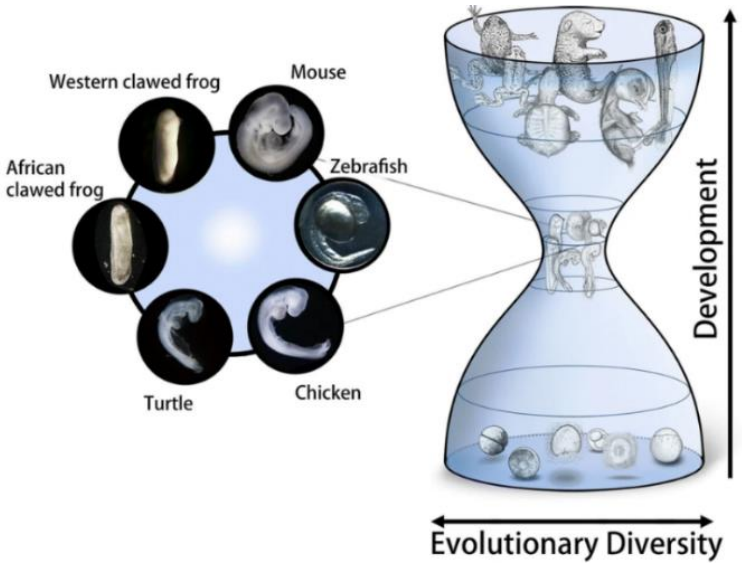


Figure 3: *The 'evo-devo hourglass model' representing embryonic development. In the midst of embryonic development, i.e. in the so-called phylotypic stage, vertebrate embryos morphologically look most similar (as opposed to earlier and later stages). During this highly conserved stage (represented by the thin part of the hourglass) embryos of all vertebrate species (example photos on the left) acquire a similar, phylum-wide body plan. It is during this stage that 50% of* Triturus *embryos succumb. [re-used image, published before by 3 under the terms of the Creative Commons Attribution 4.0 license - http://creativecommons.org/licenses/by/4.0/. The image was cropped in order to show only the photos and the hourglass visualization].*

Two hypotheses: investigating the roles of supergenes & introgression

Suppressed recombination in combination with heterozygote advantage could paves the way for the origin of a balanced lethal system [21, 63, 64]. When chiasmata are unable to form on a chromosome, for example due to chromosomal rearrangements, no physical link can be made between chromatids and cross-over of DNA is inhibited [82]. In case linkage between loci is extremely tight, supergenes comprising of multiple genes can be formed [83]. Supergenes are named as such because the linked genes are inherited together, in stretches of DNA that often do not undergo recombination and thus evolve independently of one another, which can rapidly lead to complex adaptations [82, 84-86]. Thus, supergenes may contain a set of advantageous genes and an individual can be polymorphic (i.e., can have two different versions of the supergene) if different combinations of such alleles lead to unique, advantageous outcomes – which causes balancing selection [83, 85, 87].

In case of high balancing selection and suppressed recombination, purifying selection is extremely low and deleterious mutations are accumulated rapidly by genetic drift due to Muller's Ratchett [21, 63, 85, 88], and this could continue to the point of homozygous lethality. This theory, which states that a balanced lethal system must comprise of two supergenes that got caught in a balanced polymorphism and that degraded over time due to a lack of recombination, is the working theory that was adopted by the research team of the Wielstra lab in 2019, and that forms the basis of the work presented in this dissertation: the balanced lethal system in *Triturus* posed the perfect case study to investigate this more in-depth, using modern, molecular tools. All that is necessary to figure out more about the evolution of the balanced lethal system in *Triturus* is to either use, or invent, methods that can detect the genetic differences between genomic data of heterozygote versus homozygote *Triturus* individuals, and methods that can map the differences between the DNA of *Triturus* and that of closely related, unaffected salamanders, for instance the smooth newts [of the sister genus *Lissotriton*; 89]. To summarize: all that I have to do is to, somehow, bring to light the genetic variation underlying chromosome 1 syndrome.

As mentioned before, the existence of genetic variation is paramount in nature and can be seen as the ultimate cornerstone of evolution [6, 7, 11, 15, 90, 91]. This is why it is studied so broadly, meaning that I will have (at least some) tools at my disposal to study the DNA of *Triturus*. Hybridization and introgression are currently being increasingly recognized as a source of adaptive variation in natural populations [92, 93]. In short, hybridization occurs when separated populations that are in the midst of a speciation event are still able to sexually reproduce viable offspring [94, 95]. Hence hybrids, intermediate varieties of both populations or species, are formed. In case hybrids are able to backcross with one of the parent species over many generations, introgression (also called introgressive hybridization) can take place, whereby alleles from one species are incorporated in the gene pool of the other [96, 97]. Introgression is sometimes facilitated by the act of balancing selection, which occurs as a natural result of adaptation in diploids [83, 92, 98].

Interspecific gene-flow between the *Triturus* ancestor and ancient ancestors of extant genera could have brought two uniquely adapted supergenes together into one genome. Subsequently, this could have kickstarted balancing selection in case this *Triturus* ancestor somehow benefitted from having these two forms, in turn initiating the gradual 'supergene degradation' process.

Examples of supergenes that have originated through transmission from one species into another by hybridization and introgression are becoming increasingly apparent. For example, in white-throated sparrows (*Zonotrichia albicollis*) a supergene, currently maintained in the species due to a strong, disassortative mating system, likely originated through introgressive hybridization in the past [99]. This also appears to be the

case in the Numata longwing butterfly (*Heliconius numata*), where the introgression of a divergent and inverted DNA segment has resulted in a balanced, mimicry polymorphism maintained by negative frequency-dependent selection [100]. Furthermore, rapid sex chromosome evolution constitutes an example, like the Y chromosome of the ninespine stickleback (*Pungitius pungitius*) that has been transferred from another stickleback species through ancient hybridization [101].

The hypothesis that the balanced lethal system of *Triturus* comprises two supergene variants that got caught in a balanced polymorphism by slowly degrading over time alone seems plausible, however fixation of the system is hard to explain in that case. Therefore, testing the alternative, or perhaps complementary, hypothesis that introgressive hybridization somehow played a role in bringing together chromosome 1A and chromosome 1B in *Triturus*' patient zero is not an uncanny thought. Introgressive hybridization is also not uncommon in the genus *Triturus* [102, 103] and other salamanders [104-107] and a recent study on Salamandridae supports extensive introgression at deep timescales [89]. In fact, (apart from the issue of gigantic, 'crappy' genomes) ancient episodes of hybridization that resulted in introgression are believed to be the main reason why resolving the phylogeny of salamanders has proven to be rather difficult [89, 104]. It is likely that introgressive hybridization played a causal role in the origin of chromosome 1 syndrome in *Triturus* newts, in which rapid diversification by evolutionary radiation is common [108] and in which various cases of ancient hybridization and introgression are suspected [89].

Objectives & dissertation outline

This dissertation is part of a broader research project that revolves around unraveling the evolution of balanced lethal systems. To understand why balanced lethal systems exist in nature, it is important to learn as much as possible about their genomic basis at the point of origin. To do this, the case of *Triturus* gives rise to an excellent starting point. However, as mentioned before, salamanders have gigantic genomes, which does not always make it easy to genomically study them. Thus, this dissertation comprises not only Chapters describing the most recent hypotheses and empirical insights regarding the balanced lethal system of *Triturus*, but also Chapters describing the innovative methodologies that I applied to make studying this possible in the first place.

Overall, this dissertation includes one (popular) scientific review Chapter, three research Chapters presenting new or improved methods, and a main empirical Chapter that summarizes my discoveries about the balanced lethal system of *Triturus*. As you will notice, I here highlight the key collaborators who were crucial in realizing these Chapters (in addition to the general acknowledgements), as this work was a particularly extensive

joint research effort. The Chapters are written as independent research articles and therefore they may contain some theoretical overlap. Also, the Chapters are organized chronologically to guide the reader through my scientific journey from the summer of 2019 up to the winter of 2024/2025, as summarized in Fig. 4.

CHAPTER 2 is a translated version of a Dutch (popular) scientific literature review. This Chapter focuses specifically on the main theory of the Wielstra lab that was adopted in 2019, namely that a balanced lethal system likely comprises two degraded supergene versions that got caught in a balanced polymorphism [21]. This Chapter firstly explains what a balanced lethal system is exactly, and it secondly explains in detail what the supergene theory entails. This Chapter was originally meant to be comprehensible for a non-specialist audience, and also the translated version provided here should (hopefully) interest any reader that is curious about the biology of *Triturus* newts and balanced lethal systems.

CHAPTER 3 is the first of a set of three Chapters that focus on a method that is required to start investigating the balanced lethal system in *Triturus* newts on a molecular level. Namely, to study what the differences are in the DNA of healthy versus diseased *Triturus* embryos, the initial step in the investigation is identifying which embryos are healthy (i.e. fall in the heterozygous 1A1B/1B1A class), and which ones are not (i.e., fall in one of the two diseased, homozygous classes 1A1A and 1B1B). Thus, this Chapter – *which is largely the result of a 50/50 collaboration with my co-worker Willem Meilink, and thus comprises a shared, first authorship* – introduces a molecular, laboratory procedure that can quickly tell apart the three classes of *Triturus* embryos using 1A and 1B-linked markers. This new method, which we call 'multiplex Kompetitive Allele-Specific PCR (mxKASP), is based on markers that I discovered in early tests using standard PCR methods, and it can be used to genotype *Triturus* embryos on a large scale (as classifying embryos based on morphology is a tedious and specialist task). This study explains the mxKASP method in more detail, and the broader, potential utilities of the method are described as well.

CHAPTER 4 follows naturally after Chapter 3, providing information on a target-capture method called 'NewtCap' that can be used to sequence (a part of) the DNA of *Triturus*. Namely, once it is clear to which genotypic category a *Triturus* embryo belongs (Chapter 3), the next step of the research is acquiring more in-depth genetic information about these embryos, so that the differences between the different classes (and thus, the differences between chromosomes 1A and 1B) can be mapped out and retraced to the ancestor. Moreover, this method should also be able to sequence the same (orthologous) regions of the DNA of other salamander species that do not suffer from a balanced lethal system for the sake of comparison. As whole genome sequencing of salamanders remains unattainable for now, the NewtCap tool provides an efficient solution. The study not only describes the ins and outs of the lab-protocol behind NewtCap – *which was*

initially designed by a group of the supportive co-authors, and which was impressively optimized in the laboratory by my co-worker James France – it also shows how well the method works across a wide range of species from the diverse Salamandridae family. Furthermore, instead of focusing on the usefulness of the tool for the main goal of this project (i.e., unraveling the evolution of balanced lethal systems), I emphasize that this Chapter is primarily meant to show that NewtCap in general promises to be a resourceful tool for molecular research towards Salamandridae salamanders.

CHAPTER 5 is a methodological paper that truly relies on the power of interdisciplinary research. This Chapter combines (bio)engineering and (bio)technological approaches to allow for the next step in the research process: identifying the mutations that underly chromosome 1 syndrome by searching for anomalies in the *Triturus* target capture data obtained with NewtCap (Chapter 4). The study focuses on the search for presence/absence variation, or PAV, in target capture data in general. In other words: it provides information on how to discover genes that are missing from chromosome 1A, but that are still present on chromosome 1B – and *vice versa* – using a tool called 'PAV-spotter' (validated by genotyping, see Chapter 3). Existing, bioinformatic tools were not always adequate at recognizing PAV patterns in my data due to (occasional) low coverage and low sample sizes. Thus, I reached a 'scientist's block', and for a while I ended up discussing my frustrations at home on a daily basis, over dinner. This not only unexpectedly led to the formation of the final, methodological computer-based tool required to continue my work – *something that was only possible thanks to my co-worker and husband Chris van der Ploeg, who is a software engineer* – it also led to the formation of this study. Firstly, it explains how considering the genomic position as a variable for 'signal displacement' (rather than time) enables identifying PAV in target capture data. Secondly, it provides background on potential other applications of this tool.

CHAPTER 6 comprises the main, empirical, and final research Chapter of this dissertation and focuses on the genetic background, likely origin, and evolution of the balanced lethal system in *Triturus*. Thus, from this point on the dissertation shifts from discussing theoretical frameworks and methodologies, to presenting the final research outcomes and conclusions. In a nutshell, this study compares embryos of all genotypic classes across *Triturus* species (as genotyped, see Chapter 3), based on the contents of chromosomes 1A and 1B (as sequenced, see Chapter 4), highlighting both differences and notable similarities in PAV (as determined, see Chapter 5). Additionally, the study demonstrates how detailed downstream analyses of the data at hand can uncover whether, and how, the evolutionary history of *Triturus'* chromosome 1 differs from that of the rest of the genome and that of other Salamandridae species. Lastly, the implications of all findings are described, making this Chapter the cornerstone of the dissertation.

Finally, **CHAPTER 7** concludes the dissertation with a comprehensive discussion and summary. This Chapter comprises of two parts. In the first part, the new insights into

the evolution of the balanced lethal system in *Triturus* are discussed (Chapter 2 and Chapter 6), and suggestions for further research are provided. In the second part, the various ways this dissertation contributes to both science and education are discussed, as the new insights, methods, and tools gained throughout this research extend beyond answering the specific research questions relating to chromosome 1 syndrome in *Triturus*.

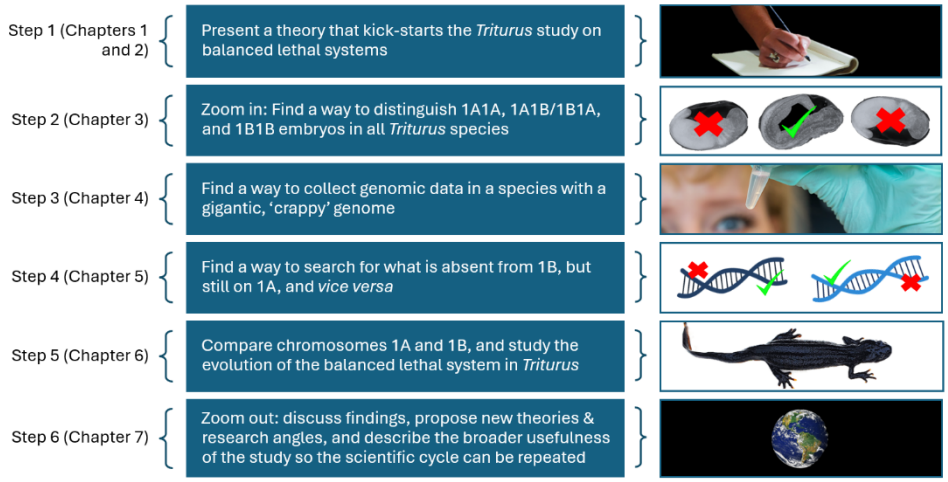


Figure 4: *A summary visualizing how the ordering of the Chapters in this dissertation corresponds to the steps undertaken in the +five year long scientific process. Explanation and sources of the photos and images: Step 1 & Step 6 – photos and images adjusted from www.pixabay.com; Step 2 – X-radia scans of* Triturus *embryos by Dr. Tijana Vučić, adjusted with symbols to indicate there are three types, namely healthy 1A1B/1B1A (checkmark) versus diseased 1A1A and 1B1B (crosses) individuals; Step 3 – photo of me in the laboratory at Leiden University, holding a test tube with freshly extracted DNA, visible as a white cloud (© Ingrid den Boer); Step 4 – representation of comparing chromosome 1A and 1B, which have unique recessive, lethal alleles, adjusted from www.pixabay.com; Step 5 – photo of a* Triturus *newt I observed in the wild during fieldwork in France (© Manon de Visser).*

References

1. Wielstra, B. and J.W. Arntzen, *Unraveling the rapid radiation of crested newts (Triturus cristatus superspecies) using complete mitogenomic sequences*. BMC Evolutionary Biology, 2011. **11**: p. 162.
2. Wielstra, B., et al., *A revised taxonomy of crested newts in the Triturus karelinii group (Amphibia: Caudata: Salamandridae), with the description of a new species*. Zootaxa, 2013. **3682**: p. 441-453.
3. Irie, N., N. Satoh, and S. Kuratani, *The phylum Vertebrata: a case for zoological recognition*. Zoological Lett, 2018. **4**: p. 32.
4. Wielstra, B., D. Salvi, and D. Canestrelli, *Genetic divergence across glacial refugia despite interglacial gene flow in a crested newt*. Evolutionary Biology, 2021. **48**(1): p. 17-26.
5. Arntzen, J.W., *Morphological and genetic diversification of pygmy and marbled newts, with the description of a new species from the wider Lisbon Peninsula (Triturus, Salamandridae)*. Contributions to Zoology, 2024. **93**(2): p. 178-200.
6. Darwin, C., *On the origin of species by means of natural selection, or preservation of favoured races in the struggle of life*. 1859.
7. Darwin, C. and A. Wallace, *On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection*. Journal of the Proceedings of the Linnean Society of London. Zoology, 1858. **3**: p. 45-62.
8. Steinberg, B. and M. Ostermeier, *Environmental changes bridge evolutionary valleys*. Science Advances, 2016. **2**: p. 1-9.
9. Wright, S., *Random drift and the shifting balance theory of evolution*, in *Mathematical topics in population genetics. Bioinformatics, vol 1*. 1970, Springer: Berlin. p. 1-31.
10. Schlaepfer, M.A., M.C. Runge, and P.W. Sherman, *Ecological and evolutionary traps*. Trends in Ecology & Evolution, 2002. **17**(10): p. 474-480.
11. Futuyma, D.J., *Evolution*. 2011.
12. Robertson, B.A. and D.T. Blumstein, *How to disarm an evolutionary trap*. Conservation Science and Practice, 2019. **1**(11).
13. Robertson, B.A. and A.D. Chalfoun, *Evolutionary traps as keys to understanding behavioral maladapation*. Current Opinion in Behavioral Sciences, 2016. **12**: p. 12-17.
14. Charlesworth, B. and D. Charlesworth, *Elements of evolutionary genetics*. 2010.
15. Frankham, R., J.D. Ballou, and D.A. Briscoe, *A primer of conservation genetics*. 1st Editio ed. 2004, New York: Cambridge University Press.
16. Mendel, G., *Versuche uber pflanzen-hybriden*. Vortlegt in den Sitzungen, 1865.
17. Lande, R., *Natural selection and random genetic drift in phenotypic evolution*. Evolution, 1976: p. 314-334.
18. Hoekstra, H.E., et al., *Strength and tempo of directional selection in the wild*. Proceedings of the National Academy of Sciences, 2001. **98**(16): p. 9157-9160.
19. García-Dorado, A., *On the consequences of ignoring purging on genetic recommendations for minimum viable population rules*. Heredity, 2015. **115**: p. 185-187.
20. Lande, R., D.W. Schemske, and S.T. Schultz, *High Inbreeding Depression, Selective Interference Among Loci, and the Threshold Selfing Rate for Purging Recessive Lethal Mutations*. Evolution, 1994. **48**: p. 965-978.
21. Wielstra, B., *Balanced lethal systems*. Current Biology, 2020. **30**: p. R742-R743.
22. Dawkins, R., *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. 1986: WW Norton & Company.
23. Lenfant, C. and K. Johansen, *Gas exchange in gill, skin, and lung breathing*. Respiration Physiology, 1972. **14**: p. 211-218.
24. Feder, M.E. and W.W. Burggren, *Skin Breathing in Vertebrates*. Scientific American, 1985. **253**: p. 126-143.
25. Towle, E.W., *On muscle regeneration in the limbs of Plethedon*. The Biological Bulletin, 1901. **2**: p. 289-299.
26. Kintner, C.R. and J.P. Brockes, *Monoclonal antibodies identify blastemal cells derived from dedifferentiating muscle in newt limb regeneration*. Nature, 1984. **308**: p. 67-69.
27. Kumar, A. and A. Simon, *Salamanders in regeneration research*. 2015.

28. Stuart, S.N., et al., *Status and trends of amphibian declines and extinctions worldwide.* Science, 2004. **306**: p. 1783-1787.
29. Araujo, A., et al., *Micro(nano)plastics as an emerging risk factor to the health of amphibian: A scientometric and systematic review.* Chemosphere, 2021. **283**: p. 131090.
30. Hof, C., et al., *Additive threats from pathogens, climate and land-use change for global amphibian diversity.* Nature, 2011. **480**(7378): p. 516-9.
31. Ford, J., et al., *Adrift on a Sea of Troubles: Can Amphibians Survive in a Human-Dominated World?1*. Herpetologica, 2020. **76**(2).
32. Préau, C., et al., *Habitat patches for newts in the face of climate change: local scale assessment combining niche modelling and graph theory.* Scientific Reports, 2020. **10**: p. 1-13.
33. Sparreboom, M., *Salamanders of the Old World: the salamanders of Europe, Asia and northern Africa*. 2014: p. 431p.
34. Van der Goot, A., M. De Visser, and A.F. Hiemastra, *Smooth newts Lissotriton vulgaris observed hibernating in a waterfowl nest.* Herpetological Bulletin, 2022. **162**: p. 41-42.
35. Glover, N.J., et al., *An experimental assessment of detection dog ability to locate great crested newts (Triturus cristatus) at distance and through soil.* PLoS ONE, 2023. **18**(6): p. e0285084.
36. Grimm-Seyfarth, A., *Environmental and training factors affect canine detection probabilities for terrestrial newt surveys*. Journal of Veterinary Behavior, 2022. **57**: p. 6-15.
37. Davic, R.D. and H.H. Welsh, *On the Ecological Roles of Salamanders*. Annual Review of Ecology, Evolution, and Systematics, 2004. **35**(1): p. 405-434.
38. Kaczmarek, J., M. Piasecka, and M. Kaczmarski, *Winter activity of the smooth newt Lissotriton vulgaris in Central Europe.* Herpetological Bulletin, 2018. **144**(2): p. 21-22.
39. O'Donnell, K.M. and R.D. Semlitsch, *Advancing Terrestrial Salamander Population Ecology: The Central Role of Imperfect Detection*. Journal of Herpetology, 2015. **49**(4): p. 533-540.
40. Gregory, T.R., *Genome size and developmental complexity.* Genetica, 2002. **115**: p. 131-146.
41. Litvinchuk, S.N., J.M. Rosanov, and L.J. Borkin, *Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the Salamandridae (Urodela, Amphibia)*. Genome, 2007. **50**: p. 333-342.
42. Gregory, T.R. *Animal Genome Size Database*. 2024 14-11-2024]; Available from: http://www.genomesize.com.
43. Sessions, S.K., *Evolutionary cytogenetics in salamanders.* Chromosome Research, 2008. **16**(1): p. 183-201.
44. Sun, C., et al., *LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders*. Genome Biology and Evolution, 2012. **4**(2): p. 168-83.
45. Sessions, S.K. and D.B. Wake, *Forever young: Linking regeneration and genome size in salamanders*. Developmental Dynamics, 2021. **250**(6): p. 768-778.
46. Rios-Carlos, H., et al., *Genomic Gigantism is not Associated with Reduced Selection Efficiency in Neotropical Salamanders.* Journal of Molecular Evolution, 2024. **92**(4): p. 371-380.
47. Wallace, H., *The balanced lethal system of crested newts.* Heredity, 1994. **73**: p. 41-46.
48. Macgregor, H.C. and H. Horner, *Heteromorphism for chromosome 1, a requirement for normal development in crested newts*. Chromosoma, 1980. **76**: p. 111-122.
49. Meilink, W.R.M., et al., *Balanced Lethal Systems: an evolutionary mystery.* Frontiers for Young Minds, 2021. **9**.
50. Horner, H.A. and H.C. Macgregor, *Normal development in newts Triturus and its arrest as a consequence of an unusual chromosomal situation*. Journal of Herpetology, 1985. **19**(2): p. 261-270.
51. D'Amen, M., L. Vignoli, and M.A. Bologna, *The normal development and the chromosome No. 1 syndrome in Triturus carnifex carnifex (Caudata, Salamandridae)*. Italian Journal of Zoology, 2006. **73**: p. 325-333.
52. Dobzhansky, T. and O. Pavlovsky, *An Extreme Case of Heterosis in a Central American Population of Drosophila tropicalis*. Proceedings of the National Academy of Sciences, 1955. **41**: p. 289-295.
53. Dawson, P.S.A., *A balanced lethal system in the flour beetle Tribolium castaneum.* Heredity, 1967. **22**: p. 435-438.
54. James, S.H., *Complex hybridity in Isotoma petraea. I. The occurrence of interchange heterozygosity, autogamy and a balanced lethal system*. Heredity, 1965. **20**: p. 341-353.
55. Steiner, E., *New aspects of the balanced lethal mechanism in Oenothera*. Genetics, 1956. **41**: p. 486-500.

56. Lin, Y.J., *Chromosome distribution and catenation in Rhoeo spathacea var. concolor.* Chromosoma, 1979. **71**: p. 109-127.
57. Thien, L.B., *Chromosome translocations in Gayophytum (Onagraceae).* Evolution, 1969. **23**: p. 456-465.
58. Rusconi, M., *Amours des salamandres aquatiques: et developpement du tetard de ces salamandres depuis l'oeuf jusqu'a l'animal parfait.* 1821.
59. Wright, S., *The roles of mutation, inbreeding, crossbreeding and selection in evolution.*, in *Sixth International Congress on Genetics*. 1932. p. 356-366.
60. Keightley, P.D. and S.P. Otto, *Interference among deleterious mutations favours sex and recombination in finite populations.* Nature, 2006. **443**: p. 89-92.
61. Ridley, M., *The Red Queen: Sex and the Evolution of Human Nature*. 1994, Penguin UK.
62. Muller, H.J., *Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors.* Genetics, 1918. **3**: p. 422.
63. Berdan, E.L., et al., *Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210199.
64. Berdan, E.L., et al., *Genomic architecture of supergenes: connecting form and function.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210192.
65. Steinfartz, S., et al., *A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary atterns of the Salamandridae with emphasis on Triturus newts.* Journal of experimental zoology. Part B, Molecular and developmental evolution, 2007. **308B**: p. 139-162.
66. Wielstra, B., *Triturus newts.* Curr Biol, 2019. **29**(4): p. R110-R111.
67. Fahrbach, M. and U. Gerlach, *The genus Triturus: History, Biology, Systematics, Captive Breeding.* 2018.
68. Sims, S.H., et al., *Chromosome 1 in crested and marbled newts (Triturus) - An extraordinary case of heteromorphism and independent chromosome evolution.* Chromosoma, 1984. **89**: p. 169-185.
69. Morgan, G.T., *Absence of chiasmata from the heteromorphic region of chromosome I during spermatogenesis in Triturus cristatus carnifex.* Chromosoma, 1978. **66**: p. 269-280.
70. Green, D.M. and S.K. Sessions, *Amphibian cytogenetics and evolution.* 2012.
71. Harrison, R.G. and S. Wilens, *Organization and Development of the Embryo.* 1969.
72. Sessions, S.K., et al., *Cytology, embryology, and evolution of the developmental arrest syndrome in newts of the genus Triturus (Caudata: Salamandridae).* The Journal of Experimental Zoology, 1988. **248**: p. 321-334.
73. Wolpert, L. and C. Tickle, *Principles of Development.* 2011: p. 616 pp.
74. Vucic, T., et al., *A staging table of Balkan crested newt embryonic development to serve as a baseline in evolutionary developmental studies.* Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 2024.
75. Irie, N. and S. Kuratani, *The developmental hourglass model: A predictor of the basic body plan?* Development, 2014. **141**: p. 4649-4655.
76. Alberts, B., et al., *Molecular Biology of the Cell: Seventh International Student Edition with Registration Card*. 2022: WW Norton & Company.
77. Irie, N. and S. Kuratani, *Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis.* Nature Communications, 2011. **2**: p. 246-248.
78. Warner, J.F., et al., *Regeneration is a partial redeployment of the embryonic gene network.* bioRxiv, 2019. **658930**.
79. Matsunami, M., et al., *A comprehensive reference transcriptome resource for the Iberian ribbed newt Pleurodeles waltl, an emerging model for developmental and regeneration biology.* DNA Research, 2019. **26**: p. 217-229.
80. Hasegawa, S., et al., *Identification and characterization of POU class V family genes in Japanese red bellied newt, Cynops pyrrhogaster.* Zygote, 2019: p. 1-8.
81. Zeller, R., J. López-Ríos, and A. Zuniga, *Vertebrate limb bud development: Moving towards integrative analysis of organogenesis.* Nature Reviews Genetics, 2009. **10**: p. 845-858.
82. Hill, W.G. and A. Robertson, *The effect of linkage on limits to artificial selection.* Genetics Research, 1966. **8**: p. 269-294.
83. Llaurens, V., A. Whibley, and M. Joron, *Genetic architecture and balancing selection: the life and death of differentiated variants.* Molecular Ecology, 2017. **26**: p. 2430-2448.

84. Schwander, T., R. Libbrecht, and L. Keller, *Supergenes and complex phenotypes*. Current Biology, 2014. **24**: p. R288-R294.
85. Pennisi, E., *'Supergenes' drive evolution*. Science, 2017. **357**: p. 1083.
86. Thompson, M.J. and C.D. Jiggins, *Supergenes and their role in evolution*. Heredity, 2014. **113**: p. 1-8.
87. Black, D. and D.M. Shuker, *Supergenes*. Current Biology, 2019. **29**(13): p. R615-R617.
88. Berdan, E.L., et al., *Muller's ratchet and the long-term fate of chromosomal inversions*. bioRxiv, 2019: p. 606012.
89. Rancilhac, L., et al., *Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders*. Molecular Phylogenetics and Evolution, 2021. **155**.
90. Väinölä, R. and K. Johannesson, *Genetic diversity and evolution*, in *Biological Oceanography of the Baltic Sea*. 2017. p. 233-253.
91. Holderegger, R., U. Kamm, and F. Gugerli, *Adaptive vs. neutral genetic diversity: Implications for landscape genetics*. Landscape Ecology, 2006. **21**: p. 797-807.
92. Fijarczyk, A., et al., *Balancing selection and introgression of newt immune-response genes*. Proceedings of the Royal Society B: Biological Sciences, 2018. **285**.
93. Li, G., et al., *Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae)*. Genome Research, 2016. **26**: p. 1-11.
94. Abbott, R., et al., *Hybridization and speciation*. Journal of Evolutionary Biology, 2013. **26**: p. 229-246.
95. Mallet, J., *Hybridization as an invasion of the genome*. Trends in Ecology and Evolution, 2005. **20**: p. 229-237.
96. Harrison, R.G. and E.L. Larson, *Hybridization, introgression, and the nature of species boundaries*. Journal of Heredity, 2014. **105**: p. 795-809.
97. Anderson, E. and L. Hubricht, *Hybridization in Tradescantia. III. the Evidence for Introgressive Hybridization*. American Journal of Botany, 1938. **25**: p. 396-402.
98. Sellis, D., et al., *Heterozygote advantage as a natural consequence of adaptation in diploids*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**: p. 20666-20671.
99. Tuttle, E.M., et al., *Divergence and functional degradation of a sex chromosome-like supergene*. Current Biology, 2016. **26**: p. 344-350.
100. Jay, P., et al., *Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion*, in *Current Biology*. 2018. p. 1839-1845.e3.
101. Dixon, G., J. Kitano, and M. Kirkpatrick, *The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes*. Molecular Biology and Evolution, 2019. **36**: p. 28-38.
102. Wielstra, B., et al., *Efficient screening for 'genetic pollution' in an anthropogenic crested newt hybrid zone*. Conservation Genetics Resources, 2016. **8**: p. 553-560.
103. Wielstra, B., et al., *A genomic footprint of hybrid zone movement in crested newts*. Evolution Letters, 2017. **1**: p. 93-101.
104. Rodríguez, A., et al., *Inferring the shallow phylogeny of true salamanders (Salamandra) by multiple phylogenomic approaches*. Molecular Phylogenetics and Evolution, 2017. **115**: p. 16-26.
105. Zieliński, P., et al., *No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (Lissotriton montandoni)*. Molecular Ecology, 2013. **22**: p. 1884-1903.
106. Canestrelli, D., R. Bisconti, and G. Nascetti, *Extensive unidirectional introgression between two salamander lineages of ancient divergence and its evolutionary implications*. Scientific Reports, 2014. **4**: p. 6516.
107. Iannella, M., P. D'Alessandro, and M. Biondi, *Evidences for a shared history for spectacled salamanders, haplotypes and climate*. Scientific Reports, 2018. **8**(1): p. 16507.
108. Wielstra, B., et al., *Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle*. Molecular Phylogenetics and Evolution, 2019. **133**: p. 120-127.

Freshly laid Triturus *egg in the wild*

Chapter 2 - An evolutionary mystery: the deadly chromosome 1 syndrome in *Triturus* salamanders

Manon de Visser[1,2], James France[1,2], Willem Meilink[1,2] and Ben Wielstra[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands
2. Understanding Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands

Originally, this popular science article is published as:

De Visser, M.C., France, J.M., Meilink, W.R.M. & Wielstra, B.M. (2021). Een evolutionair raadsel: het dodelijke chromosoom 1 syndroom in *Triturus*-salamanders, ***RAVON*** 23(1): 9-12.

For this Chapter, the Dutch article has been freely translated to English. The lead text has been left out, and the English summary that was already in the article provided at the end is here provided as the abstract.

Abstract

Crested and marbled newts (genus: *Triturus*) are stuck with a the deadly 'chromosome 1 syndrome', which is caused by lethal, genetic mutations. The syndrome is a result of a so-called balanced lethal system. Normally, natural selection will suppress lethal mutations. However, in a balanced lethal system unique, recessive lethal mutations exist on two different versions of a chromosome. These two versions compensate for one another: a functional gene copy on one chromosome type masks the damaged gene present on the other chromosome type, and vice versa. This way, the two different types of chromosome are both required to survive. Because all homozygotes die, this leads to a reproductive output that is cut in half, each generation. The *Triturus* newts are not the only organisms that suffer from such a wasteful system, as similar syndromes have been described in insects and plants. This is an evolutionary paradox: why would something so disadvantageous evolve time and again? Modern DNA-sequencing techniques allow us to compare *Triturus* DNA with that from newt species that do not suffer from chromosome 1 syndrome, to try and decipher the genetic basis of the syndrome. This will help us understand how balanced lethal systems repeatedly evolve in nature by the 'hijacking' of standard evolutionary processes.

Introduction

According to evolutionary theory, natural selection reduces the likelihood that a lethal mutation that disrupts a crucial gene will be passed on to the next generation [1]. In this way, genetically determined diseases are suppressed in natural populations. However, newts of the genus *Triturus* are affected by the so-called 'chromosome 1 syndrome' [2] which causes 50% of the eggs to not hatch (Figure 1). All *Triturus* species, i.e. seven crested newt species and two marbled newt species, suffer from chromosome 1 syndrome [3].

The phenomenon is an evolutionary mystery that has puzzled scientists for two centuries. Mauro Rusconi was the first to realize that half of the eggs of crested newts stop growing halfway through their development. In his now two-hundred-year-old book *'Amours des Salamandres Aquatiques'* ('The Love Life of Newts'), he wrote the following regarding the relevant stages of embryonic development (freely translated into English): "Although this period is most entertaining for a naturalist to study, it appears that this stage is dangerous for the little embryos themselves, as almost half of them die at this moment - or shortly after" [4].

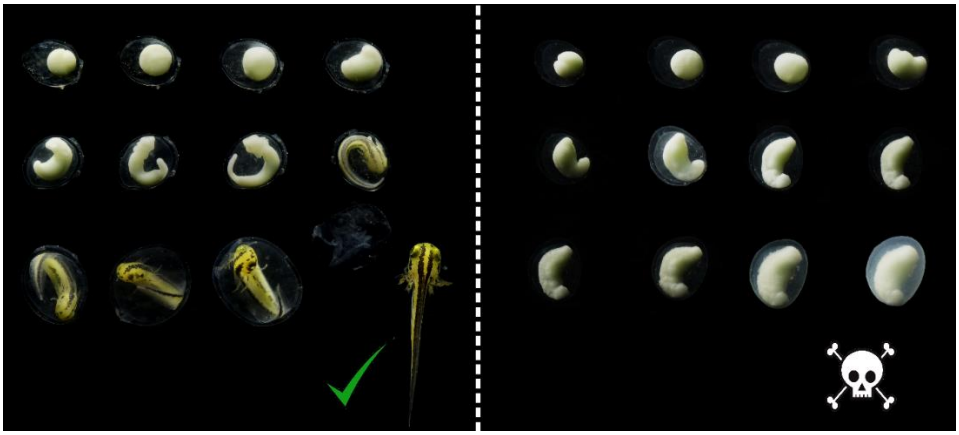


Figure 1: *Half of the eggs laid by crested and marbled newts (genus:* Triturus*) do not hatch. Out of all eggs 50% go through embryonic development successfully (left), while the other 50% of eggs stop growing halfway through embryonic development and eventually succumb (right). (Photo's: Michael Fahrbach).*

Balanced lethal system

At the basis of chromosome 1 syndrome lies a so-called 'balanced lethal system' (Wielstra, 2020). Chromosome 1 is the longest pair of the 12 chromosome pairs that *Triturus* species possess. While a normal chromosome pair consists of two very similar

versions of a chromosome, there are to be two very different versions of chromosome 1 in *Triturus* newts [5]. An embryo needs both versions of chromosome 1, also known as chromosome 1A and 1B, in order to survive. They must thus be heterozygous for chromosome 1. Homozygotes, that have either two 1A or two 1B versions, die [2]. The genetic mechanism likely works as follows: both versions of chromosome 1 contain unique, harmful mutations in crucial genes that are not present on the other version. There are thus two so-called 'alleles' (gene variants) of these crucial genes: a lethal, non-functioning allele and a normal, unaffected allele. The lethal mutations on the 1A version of chromosome 1 can be compensated for by non-mutated, still functional genes on the 1B version, and *vice versa*. Hence, the defective genes and the genes that are still functional balance each other out. This is where the term 'balanced lethal system' comes from. In short: as long as an embryo possesses both versions of chromosome 1, so one 1A and one 1B (i.e., as long as an embryo is heterozygous), the lethal alleles are suppressed. However, if an embryo has two of the same versions, so either two times 1A or two times 1B (i.e., if an embryo is heterozygous), then things go wrong.

Because only heterozygotes survive, there is a consistent loss of 50% of the eggs. This follows directly from Mendel's laws of inheritance. Namely, when two heterozygous individuals with the genotype '1A1B' or '1B1A' mate, this results in 50% of the offspring being heterozygous. The remaining 50% of individuals are homozygous: 25% end up with the genotype '1A1A' and 25% with '1B1B' (Figure 2). The heterozygotes are lucky and get a chance to grow into adults and reproduce, while both types of homozygotes are doomed to die in the egg [2]. And every generation, this macabre cycle starts all over. What a waste!

Originated millions of years ago

All *Triturus* species are affected by the chromosome 1 syndrome, while newts of the most related genus *Lissotriton* are not (for instance, the smooth newt *L. vulgaris* or the palmate newt *L. helveticus*, which are both found in the Netherlands). Therefore, it is likely that this anomaly originated in a common ancestor of *Triturus*. This means this system “got stuck” in the DNA of *Triturus* at least 24 million years ago [6, 7].

In scientific literature, chromosome 1 syndrome is *the* classic example of a balanced lethal system [8]. However, in addition to the newts there are examples of other organisms displaying a similar phenomenon, for instance some insects and plants [9]. In those cases, it is also the presence of two versions of a certain chromosome that plays a key role. The fact that balanced lethal systems occur in a broad range of taxonomic groups makes it plausible that; 1) their repeated emergence must be explainable on the basis of evolutionary principles, and 2) their occurrence may not be as rare as is currently believed. But then how could such a balanced lethal system actually arise?

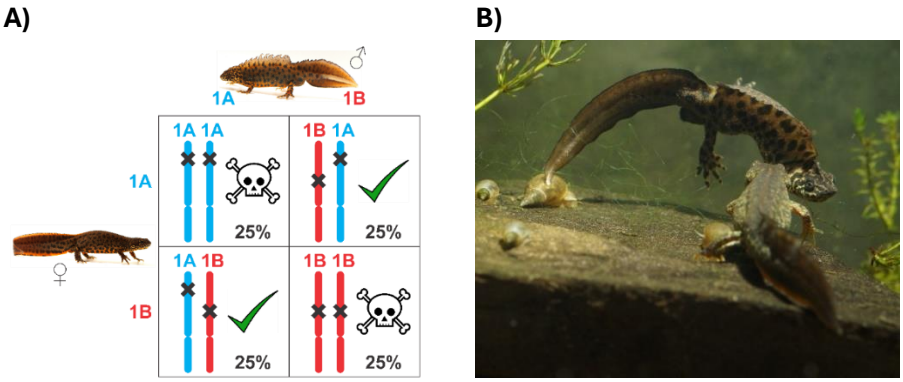


Figure 2: *Crested and marbled newts suffer from a balanced lethal system, in which only heterozygotes survive.* ***A)*** *The two different versions of chromosome 1 possess unique, lethal mutations (crosses). Adult individuals always have a 1A and a 1B version and pass on one of the two to a sex cell with a 50% chance. After fertilization four combinations are possible, each with a chance of 50% x 50% = 25%. The heterozygotes survive because of their lethal mutations are compensated for, while both types of homozygotes are doomed to die.* ***B)*** *From the mating ritual (displayed on the photo) all the way through egg deposition:* Triturus' *chromosome 1 syndrome is considered an evolutionary mystery, because of the substantial amount of energy used in the reproductive process. (Photo: Michael Fahrbach)*

Suppressed recombination

In regular chromosome pairs 'recombination' occurs during the production of sex cells (egg and sperm cells). Within a chromosome pair, this allows a part of one chromosome to be exchanged with the equivalent part of the other chromosome, and the other way around. This exchange is also called crossing-over and, in newts, it occurs on average twice on one chromosome per each cell division. Crossing-over is essential in evolution, as it ensures the constant creation of chromosomes that have a unique DNA code. Chromosome pairs are then split – a process called 'segregation' – and, one chromosome per pair is passed on to each sex cell. Thus, recombination ensures that unique combinations of chromosomes ultimately come together in the offspring during fertilization (Figure 3).

Crossing-over could theoretically swap the non-functional alleles on one of the two versions of chromosome 1 for the functional alleles from the other version. In this way, chromosome 1 syndrome could disappear in one go, as individuals with a 'corrected' chromosome 1 would suddenly be able to produce twice as many offspring: a huge selective advantage! However, crossing-over does not occur over a large part of chromosome 1 in *Triturus* [10, 11]. This suggests that the difference between chromosome 1A and 1B is probably so large that they no longer recognize each other as

equivalents, making crossing-over impossible. The result is that each sex cell possesses either chromosome 1A or 1B, but never a mix of both. Thus, if an egg and a sperm cell that carry the same version of chromosome 1 come together, this eventually results in a non-viable individual (Figure 3).

From ‘supergene’ to ‘bloopergene’

Suppressed crossing-over is often observed in so-called 'supergenes.' Supergenes essentially comprise of a group of genes that lie on a piece of a chromosome and form alleles that are inherited together, i.e. as one overall stretch. In other words: these genes are 'linked' to each other. There are always at least two versions of a supergene, each with their own set of alleles for the genes involved. This allows the alleles on each supergene to evolve together, which makes it possible for complex adaptations to arise, hence the name: 'supergene.' Individuals heterozygous for a particular supergene may have a selective advantage over individuals that are homozygous. In that case, a so-called 'balancing selection’ exists, which leads to the retention of both the supergene versions across generations [12].

Supergenes are super, but they also have a downside. Because crossing-over between different supergene variants cannot occur within heterozygotes, it is difficult to eliminate any harmful mutations that arise. Thus, what follows is a slow accumulation of harmful mutations. In the field of evolutionary genetics, this phenomenon is referred to as 'Muller's Ratchet.' In heterozygotes, a still functioning allele on the other supergene variant can still compensate for such a new harmful mutation. But if both supergenes accumulate unique lethal alleles, no homozygote will be viable. In other words: a balanced lethal system, where all homozygotes are doomed to die, has originated.

It is thus likely that the supergenes initially provided an evolutionary advantage, but that the accumulation of harmful mutations spiraled out of control in such a way that this resulted in an irreversible, deadly syndrome. And once a balanced lethal system is fixed within a species, there is no escaping it. In this case, there must have been some sort of tipping point where 'supergenes' turned into what can be considered 'blooper genes.' In a way, it is reminiscent of the foolish greediness of King Midas from Greek mythology, who chose to have everything he touched turn to gold. This seemed like a good idea in the short term, but it turned out to be disastrous in the long term.

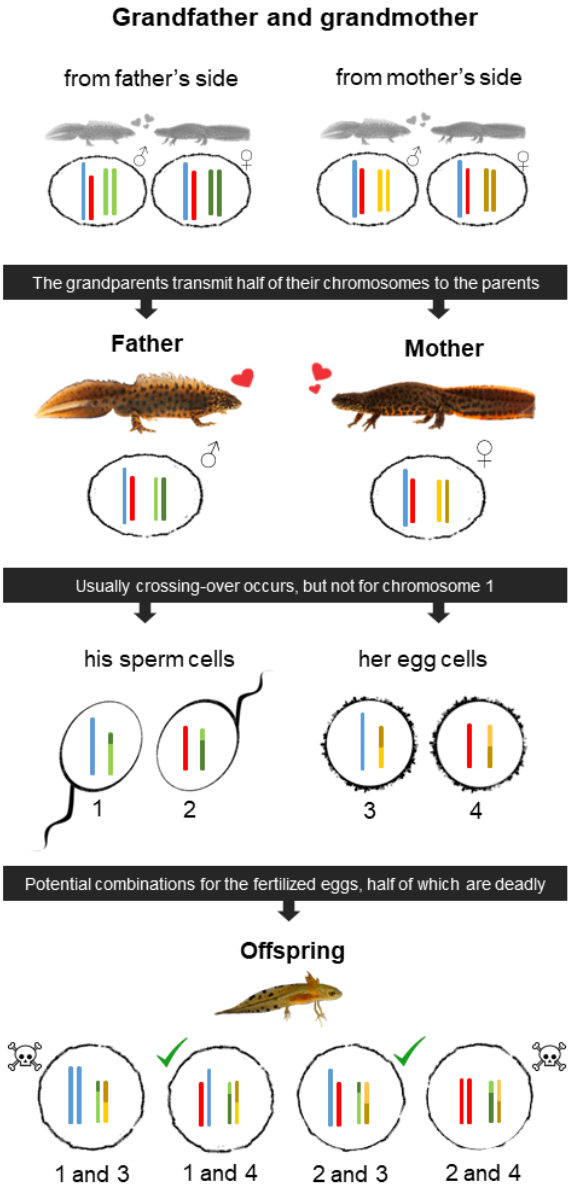


Figure 3: *Chromosome pairs usually shuffle and exchange DNA, however chromosome 1 in the genus* Triturus *poses an exception. When fathers and mothers produce sperm and egg cells, it is generally the case that crossing-over happens in the chromosomes of the grandparents (shown for one chromosome pair, displayed with two shades of green or yellow for the grandparents of father's and mother's side, respectively). However, between the two versions of chromosome 1, 1A (in red) and 1B (in blue) there is no crossing-over, meaning that an egg cell or a sperm cell can receive 1A or 1B, but never a mix of both. (Photo's: Michael Fahrbach).*

DNA research will solve the puzzle

So far, scientists have only been able to look at the overall development of *Triturus* embryos in the egg (Figure 4) and at the 'karyotype' (the general shape of the chromosomes) of these newts. Investigating it more in depth by means of DNA research used to be virtually impossible – until now. Modern DNA sequencing techniques should make it possible to identify the genes responsible for the catastrophe that befalls half the *Triturus* eggs.

By comparing the genome of *Triturus* species with that of other salamander species that do not suffer from the chromosome 1 syndrome, such as *Lissotriton* species, the lethal mutations hidden in the DNA of *Triturus* can be discovered. Furthermore, by following the line of evolutionary descent it is possible to "reconstruct" what chromosome 1 must have looked like in the ancestor of all the crested and the marbled newts, so before the syndrome originated.

Figuring out how chromosome 1 syndrome evolved hopefully leads to an understanding of how a balanced lethal system can be created by standard evolutionary processes, despite it *seeming* evolutionarily impossible.

***Figure 4**: A viable* Triturus *embryo inside the egg. (Photo: Michael Fahrbach).*

References

1. Darwin, C., *On the origin of species by means of natural selection, or preservation of favoured races in the struggle of life*. 1859.
2. Macgregor, H.C. and H. Horner, *Heteromorphism for chromosome 1, a requirement for normal development in crested newts.* Chromosoma, 1980. **76**: p. 111-122.
3. Fahrbach, M. and U. Gerlach, *The genus Triturus: History, Biology, Systematics, Captive Breeding.* 2018.
4. Rusconi, M., *Amours des salamandres aquatiques: et developpement du tetard de ces salamandres depuis l'oeuf jusqu'a l'animal parfait.* 1821.
5. Callan, H.G. and L. Lloyd, *Lampbrush Chromosomes of Crested Newts Triturus cristatus (Laurenti).* Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1960. **243**: p. 135-219.
6. Steinfartz, S., et al., *A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary atterns of the Salamandridae with emphasis on Triturus newts.* Journal of experimental zoology. Part B, Molecular and developmental evolution, 2007. **308B**: p. 139-162.
7. Rancilhac, L., et al., *Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders.* Molecular Phylogenetics and Evolution, 2021. **155**.
8. Wielstra, B., *Balanced lethal systems.* Current Biology, 2020. **30**: p. R742-R743.
9. Grossen, C., S. Neuenschwander, and N. Perrin, *The balanced lethal system of crested newts: A ghost of sex chromosomes past?* American Naturalist, 2012. **180**: p. E174-E183.
10. Wickbom, T., *Cytological Studies on Dipnoi, Urodela, Anura and Emys.* Hereditas, 1945. **31**: p. 241-346.
11. White, M.J.D., *The spermatogenesis of hybrids between Triturus cristatus and T. marmoratus (Urodela).* The Journal of Experimental Zoology, 1946. **102**: p. 179-207.
12. Schwander, T., R. Libbrecht, and L. Keller, *Supergenes and complex phenotypes.* Current Biology, 2014. **24**: p. R288-R294.

Triturus egg being prepped for DNA extraction

Chapter 3 - Determining zygosity with multiplex Kompetitive Allele-Specific PCR (mxKASP) genotyping

Manon C. de Visser[1,2*], Willem R.M. Meilink[1,2*], Anagnostis Theodoropoulos[1,2], Michael Fahrbach[3], Ben Wielstra[1,2]

1. Institute of Biology Leiden, Leiden University, Leiden, The Netherlands
2. Naturalis Biodiversity Center, Leiden, The Netherlands
3. Niedernhaller Str. 8/ 2, 74653 Criesbach, Germany

* = *These authors contributed equally to this work*

This manuscript has been submitted to the journal ***Molecular Ecology Resources*** and is currently under review. A preprinted version of this manuscript has been published at ***bioRxiv*** (2024.10.25.620256; https://doi.org/10.1101/2024.10.25.620256).

Abstract

We introduce multiplex Kompetitive Allele-Specific PCR (mxKASP): a modification of 'classical' KASP genotyping that allows zygosity to be determined in diploid organisms. Rather than targeting a SNP associated with a single marker, mxKASP targets two non-homologous markers. We show proof of concept by applying mxKASP to the balanced lethal system in *Triturus* newts, in which individuals are known to possess either: (1) zero copies of the 1A version of chromosome 1 and two copies of the 1B version; (2) one copy of 1A and one copy of 1B; or (3) two copies of 1A and zero copies of 1B. mxKASP is successful in amplifying both a 1A and a 1B marker in a single reaction (if present), allowing the zygosity of individuals to be inferred. We independently confirm our mxKASP results with a multiplex PCR approach. We argue that mxKASP can be applied to rapidly and economically determine zygostity in diploid organisms, for a large number of samples at once.

Keywords

Balanced lethal system; Multiplex PCR; *Triturus*; Sex chromosome; Supergene

Introduction

In diploid organisms, a stretch of DNA, such as a gene or even an entire chromosome, is *hemizygous* if only a single copy is present in the genome. The enormous impact (hemi)zygosity can have on the phenotype is perhaps best showcased by *sex chromosomes*. In species with genetic sex determination, the heterogametic sex is hemizygous for the region that underpins sex, whereas the homogametic sex is homozygous [1]. Zygosity also varies in the case of *supergenes*, which are physically linked genes that are inherited together because recombination is suppressed [2-4]. In fact, sex chromosomes are generally considered a kind of supergene [4]. Supergenes come in (at least) two versions and which version(s) an individual possesses – two different versions or the same version twice – may greatly influence its phenotype. Hence, zygosity is of great biological relevance and techniques that allow zygosity to be determined would be widely applicable, given the widespread occurrence of sex chromosomes and other supergenes across the tree of life. Ideally, such a technique would be highly capable of processing many samples simultaneously and rapidly at low cost.

Kompetitive Allele Specific PCR (KASP) is an efficient and cost-effective method for genotyping SNPs in a large number of samples [5]. In this method, typically a single common reverse primer is used in combination with two allele-specific forward primers that differ in the last nucleotide, defined by the targeted SNP. These primers also possess a tail that is complementary to one of the two quenched fluor-labelled oligos; one with FAM, one with HEX. When thermal cycling exponentially increases one tail (in individuals homozygous for one or the other SNP) or both tails (in heterozygous individuals), the complementary label is no longer quenched, resulting in a corresponding fluorescent signal. However, with a slight modification, KASP could, in theory, also be used to efficiently genotype markers for zygosity. Rather than targeting a single marker with SNP-specific primers, as in 'classical' KASP, two different (i.e. non-homologous) markers that are positioned on the genomic region relevant for zygosity could be targeted, to determine the presence or absence of these markers. To this aim we use two distinct primer pairs, each with a different quenched fluor-labelled oligo, in a single reaction (Fig. 1). We call this modification multiplex KASP (mxKASP).

We show proof of concept by applying mxKASP in a balanced lethal system. In a balanced lethal system two versions of a chromosome are involved [6-8]. Healthy individuals possess a single copy of both chromosome versions, meaning they are reciprocally hemizygous, while diseased individuals possess two identical chromosome versions, meaning they are homozygous for one or the other version. The best studied example of a balanced lethal system concerns the *chromosome 1 syndrome* in newts of

the genus *Triturus*, in which two versions of chromosome 1 are involved: 1A and 1B [9-11]. Only individuals that possess both versions of chromosome 1 (1A1B) survive embryonic development, whereas individuals that have the same version twice (1A1A or 1B1B), and therefore lack the alternate version, perish inside the egg. We design KASP markers that are positioned either on 1A or 1B, with the expectation that only one or the other marker would be amplified in diseased (1A1A or 1B1B) individuals, whereas both should amplify in healthy (1A1B) individuals.

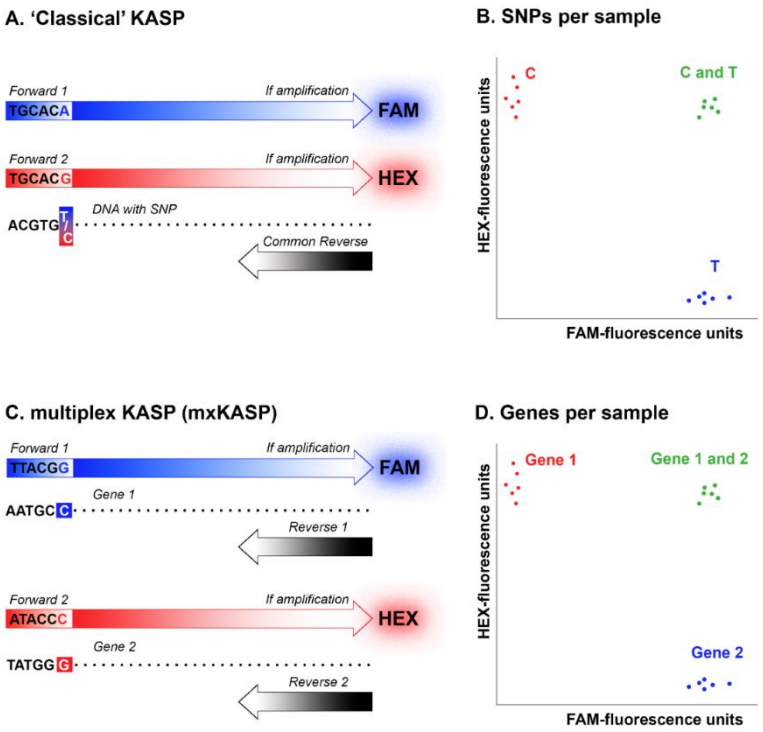


Figure 1. *Schematic showing the different approach between KASP and mxKASP. (A) In KASP two forward primers differ in their last nucleotide as defined by the targeted SNP. Depending on which SNP variant is present, amplification leads to a fluorescent signal of FAM, HEX, or both. (B) For individuals homozygous for one SNP variant (in this case C) only the HEX signal is amplified (red dots), while for individuals homozygous for the other SNP variant (T) only the FAM signal is amplified (blue dots); for heterozygous individuals (C/T) both fluorescent signals are emitted (green dots). (C) In mxKASP two pairs of forward and reverse primers are designed for different (non-homologus) genes. (D) The presence of only one or the other gene will result in either the HEX (in this case for gene 1, red dots) or the FAM (gene 2, blue dots) signal to be amplified, while if both genes are present, both fluorescent signals are emitted (green dots).*

Material and Methods

Sample selection

Embryos representing nine *Triturus* species [12], including the two distinct Balkan and Italian clades of *T. carnifex* [13] were included in the study. We included three individuals per genotype (1A1B, 1A1A, and 1B1B), adding up to 90 *Triturus* individuals in total. We also included three individuals of *Lissotriton vulgaris*, a species that is closely related to *Triturus* [14], but that is not affected by the balanced lethal system [15].

Marker selection

[16] produced DNA sequence data for c. 7k nuclear DNA markers from across the genome of three healthy (1A1B) and six diseased (three 1A1A and three 1B1B) individuals for the same ten *Triturus* clades as above (i.e. 90 samples in total). This dataset includes dozens of 1A-linked markers (not present in 1B1B individuals) and 1B-linked markers (not present in 1A1A individuals), next to thousands of additional markers positioned elsewhere on the genome. We also included three individuals of *Lissotriton vulgaris* from [17]. We took sequences of the (in *Triturus*) 1A-linked marker *PLEKHM1*, the 1B-linked marker *NAGLU* and the 'control marker' *CDK17*, found on another linkage group than the 1A and 1B marker [18, 19]. Sequences were aligned and trimmed in MEGA 11 [20], using the Muscle alignment tool set. Next, Unipro UGENE v.33 [21] was used to generate consensus sequences, with polymorphisms encoded by the relevant IUPAC codes. The consensus sequences were generated using an 80% threshold value.

mxKASP

Four primer pairs were designed for the 1A-linked marker *PLEKHM1* and an additional four for the 1B-linked marker *NAGLU*, using LGC genomics' Kraken v.23.11. This resulted in 16 (four by four) possible primer pair combinations (Supplementary Table S1, see Zenodo). KASP assays were prepared in a total volume of 200 µl, with 24 µl of both forward primers (10 µM), 60 µl of both reverse primers (10 µM), and 32 µl purified water. Next, KASP mix was made for each primer combination and contained 1.26 µl KASP assay, 102.8 µl KASP master mix and 102.8 µl purified water. All 16 combinations were tested in a 1,536 well plate with 1 µl of KASP mix and 1 µl of DNA for 93 samples (90 *Triturus*, 3 *Lissotriton*), as well as a negative control (no DNA). mxKASP was performed in a hydrocycler, beginning with 10 cycles with an annealing temperature decreasing from 61 to 55°C, followed by 26 cycles at 55°C. Two additional three-cycle recyclings were performed with annealing at 57°C. Fluorescence was measured using a PHERAstar plate reader and data were analysed using Kraken.

Two primer pair combinations (1A-linked primer pairs NAGLU.2 or NAGLU.3 with 1B-linked marker primer pair PLEKHM1.3; Table 1) showed separation into three clear groups (see results), corresponding to the three potential genotypes (1A1A, 1A1B, 1B1B). These two combinations were repeated in a 384 well plate to confirm the results. The KASP mix for each primer pair consisted of 7.2 µl assay, 194.4 µl KASP master mix and 194.4 µl purified water. Here, 3 µl of DNA of the same samples, next to a negative control, were combined with 3 µl of KASP mix, with identical downstream procedures as before.

mxPCR

To confirm our mxKASP results we designed a multiplex PCR (mxPCR) protocol. Primers were developed in Primer3Plus [22], aiming for a size difference (to be able to distinguish markers by gel electrophoresis) between the 1A-linked marker *PLEKHM1* (200-300bp), the 1B-linked marker *NAGLU* (350-450bp) and the control marker CDK17 (>500bp). All other parameters were kept the same (Table 1). The mxPCR was performed using QIAgen multiplex PCR master mix. PCR was performed in 12 µl reactions, containing 0.06 µl of all forward and reverse primers (at 10 µM concentration, resulting in a 0.05 µM end concentration of each primer), 6 µl QIAGEN multiplex PCR master mix, 4.64 µl purified water and 1 µl of DNA. PCR conditions were as follows; a hot start for 15 minutes at 95°C, followed by 35 cycles of denaturation for 30 seconds at 95°C, annealing for 1 minute at 56°C, and extension for 1 minute at 72°C, and a final ten-minute extension at 72°C.

Table 1. List of primers used for mxKASP (top six) and mxPCR (bottom six).

Chromosome	Marker	Primer	Sequence	Dye
1A	PLEKHM1	PLEKHM1.3_F	#AGTTCTCAAAGCAGGTCCGGTC	FAM
1A	PLEKHM1	PLEKHM1.3_R	GAGTTGTCTCCATTTGAGTTACGCC	NA
1B	NAGLU	NAGLU.2_F	*AGCTTGGTTGAACTCATCCTGGT	HEX
1B	NAGLU	NAGLU.3_F	*GCTAGTGGCTAAGGATTGAGCAT	HEX
1B	NAGLU	NAGLU.2_R	ACRCTAGTGGAATGCCTCAACACG	NA
1B	NAGLU	NAGLU.3_R	CACTTCCTACTTGGCAGTTGGCTTT	NA
Chromosome	**Marker**	**Primer**	**Sequence**	**Product Length (bp)**
1A	PLEKHM1	PLEKHM1_Fm	CCAGTGGCATTTTCAGGTCC	336
1A	PLEKHM1	PLEKHM1_R3	GGGAGTTCTTCTGCGAGTTG	
1B	NAGLU	NAGLU_F3	TGGTTGAACTCATCCTGGTG	416
1B	NAGLU	NAGLU_R3	GTTGATGTCACAAGGCAAGC	
Control	CDK17	CDK17_F1	GGCATGGGAAGAACAGAAGA	537
Control	CDK17	CDK17_R1	CCATCTGCTTGGACTGTTGA	

*# (binds to FAM) and * (binds to HEX) indicate the location of the LGC-patented dye-complementary tail sequence.*

Results

In our mxKASP experiment, two out of sixteen combinations of primer pairs resulted in a clear separation of three genotype clusters of the samples (Fig. 2; Supplementary Table S2, see Zenodo). The combination PLEKHM1.3 and NAGLU.2 showed allocation of all individuals into the expected genotype clusters, while two 1A1B samples were outliers for the combination of primers PLEKHM1.3 and NAGLU.3 (Fig. 2). The mxPCR amplification genotyped each individual as expected (Fig. 3).

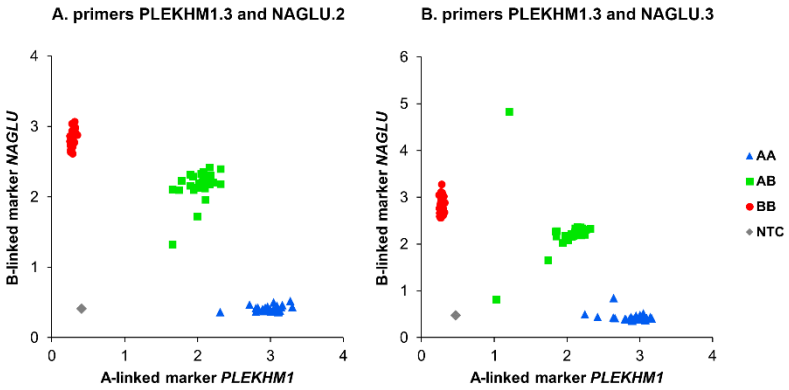


Figure 2. *KASP results for two markers on different chromosomes, one on chromosome 1A (*PLEKHM1*) and one on chromosome 1B (*NAGLU*). Primers PLEKHM1.3 and NAGLU.2 (A. left panel) lead to complete separation of the samples into three clusters, while primers PLEKHM1.3 and NAGLU.3 (B. right panel) show a similar division with outliers for two 1A1B individuals. The x-axis reflects FAM and the y-axis HEX fluorescence units (y-axes are not to scale).*

Discussion

Our modification of KASP, mxKASP, proved to be successful in determining zygosity. Instead of targeting different SNPs on a single marker, as is typically done with KASP, we show that it is possible to target completely different, non-homologous markers. We use mxKASP to independently establish the presence or absence of two markers, which allows us to infer zygosity in the *Triturus* balanced lethal system. 'Classical' KASP has been used before to determine zygosity by targeting a SNP distinguishing the X and Y chromosomes, that could be identified because it co-segregated with the male-specific region [23]. Our approach does not require such a SNP discovery step; rather it would directly target a marker private to the male-specific region, next to a marker present in both males and females.

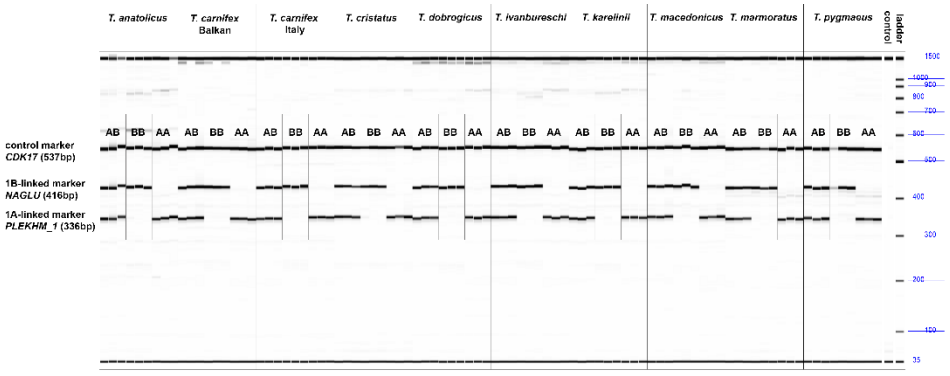


***Figure 3**. Multiplex PCR results obtained with a Fragment Analyzer. Three genotypes (healthy 1A1B or diseased 1A1A or 1B1B) can be distinguished for ten* Triturus *(candidate) species.*

Earlier KASP modifications could potentially inform on zygosity. 'Duplex KASP (dKASP)' was shown to be able to amplify two SNPs on two paralogous genes in a single run, with one of the two potential SNP variants for each paralog being present, resulting in a fluorescent signal [24]. This could be applicable to non-homologous genes as well, but depends on the presence of SNPs, which is not the case for mxKASP. The KASP modification '4-fluorescent KASP' amplifies two different SNPs, positioned on two non-homologous markers, in a single run, by using four rather than the standard two fluorescent cassettes [25]. Our mxKASP approach is simpler because it does not require additional fluorescent cassettes and is not dependent on the presence of SNPs on the markers of interest.

Over the last two decades KASP has become increasingly popular as a uniplex SNP genotyping platform for the agricultural, biological, forensic, and medical sciences [26-28]. There are several reasons as to why it is applied so broadly. Firstly, KASP is capable of processing large sample sizes: it can be run in 96, 384 or 1,536 well-plates, with up to fourteen plates per PCR program [5]. Secondly, genotyping data interpretation in Kraken is practically automated, as well as instantaneous. In comparison, the capacity of general uni- or multiplex PCR approaches is limited by the number of available PCR machines and data interpretation is relatively time consuming, relying on labor intensive gel or capillary electrophoresis. Therefore, when sample sizes are large, KASP (including modifications such as mxKASP) is more efficient.

Acquiring clean-cut KASP results depends – as is the case for any genotyping method – on the quality of input DNA [29-32]. However, KASP only requires low quantity and quality DNA. Therefore, if KASP is combined with quick and cheap upstream methods, this opens up possibilities for scientists to upscale their research to the point of being able to sample, extract and genotype thousands of individuals in a matter of days

[33]. For this reason, KASP has already been embraced as a more economical choice for quality control analysis in for instance plant breeding and seed systems, rather than genotyping by sequencing [5, 34]. The same applies to the clinical and medical fields, which incorporate KASP to rapidly and accurately identify blood groups for blood donors [35], as well as for screening for genetic susceptibility to diseases such as Parkinson's and thrombosis [36, 37].

We foresee multiple applications in which our KASP modification, mxKASP, can be used for large-scale zygosity level determination. The method can be used to improve animal welfare. For instance, mxKASP could replace capillary electrophoresis of PCR products or real-time quantitative PCR previously recommended for use in the poultry industry to sex embryos *in ovo* to reduce male chick culling [38]. Any PCR-based method would avoid the controversial approach of genetically modifying layer hens for determining the sex of their eggs [39]. mxKASP could also prove valuable in a conservation context, for example, in species where sex is influenced by both a genetic and a thermal component [40, 41]. Circa 1/3 of fish species thought to exhibit only temperature-dependent sex determination appears to possess genetic sex determination too, and which one predominates depends on environmental factors [42]. Global warming threatens such species by biasing sex ratios. As more sex-linked markers are discovered, more genotype/phenotype mismatches will presumably be revealed [41, 43]. Finally, mxKASP can be used to detect zygosity in fundamental eco-evolutionary studies of supergene systems. For example, to test the ratios of different genotypes when a particular supergene version is assumed to carry recessive deleterious mutations [44, 45].

Acknowledgements

This project received funding from the Dutch Research Council (NWO Promotiebeurs voor leraren 023.016.006) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 802759). Salamander embryos were obtained by MF from his personal breeding colony. Housing and breeding salamanders as a private individual are not considered to require a license (BGBl. I S. 3125, 3126, 3750). Housing and breeding protocols comply to EU directive standards (EU directive annex III, section B, Table 9.1) and were reviewed by the Animal Welfare Body Leiden. As per EU legislation regarding the protection of animals used for scientific purposes (EU directive no. 2010/63/EU), sacrificing embryos that are not feeding independently does not qualify as an animal experiment. Carola Feijt, James France, Bobbie Sewalt and Klaas Vrieling helped with laboratory work.

Conflict of interest

The authors declare no conflicts of interest. The funding bodies had no direct role in the design of the study nor in the collection, analysis, interpretation of data or in the writing of the manuscript.

Data accessibility and Benefit-sharing statement

All generated data have been published with the *bioRxiv* preprint and can be found in the supplementary tables online via Zenodo.

Author contributions

W.R.M.M., M.C.dV. and B.W. conceived the research. W.R.M.M., M.C.dV and A.T. conducted labwork. W.R.M.M. and M.C.dV analyzed the data. M.F. provided samples. W.R.M.M., M.C.dV. and B.W. wrote the manuscript with input from the other authors.

References

1. Bachtrog, D., et al., *Sex determination: why so many ways of doing it?* PLoS Biology, 2014. **12**(7): p. e1001899.
2. Thompson, M.J. and C.D. Jiggins, *Supergenes and their role in evolution.* Heredity, 2014. **113**: p. 1-8.
3. Berdan, E.L., et al., *Genomic architecture of supergenes: connecting form and function.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210192.
4. Schwander, T., R. Libbrecht, and L. Keller, *Supergenes and complex phenotypes.* Current Biology, 2014. **24**: p. R288-R294.
5. Semagn, K., et al., *Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement.* Molecular Breeding, 2013. **33**(1): p. 1-14.
6. Berdan, E.L., et al., *Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210199.
7. Wielstra, B., *Balanced lethal systems.* Current Biology, 2020. **30**: p. R742-R743.
8. Muller, H.J., *Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors.* Genetics, 1918. **3**: p. 422.
9. Sessions, S.K., et al., *Cytology, embryology, and evolution of the developmental arrest syndrome in newts of the genus Triturus (Caudata: Salamandridae).* The Journal of Experimental Zoology, 1988. **248**: p. 321-334.
10. Sims, S.H., et al., *Chromosome 1 in crested and marbled newts (Triturus) - An extraordinary case of heteromorphism and independent chromosome evolution.* Chromosoma, 1984. **89**: p. 169-185.
11. Macgregor, H.C. and H. Horner, *Heteromorphism for chromosome 1, a requirement for normal development in crested newts.* Chromosoma, 1980. **76**: p. 111-122.
12. Wielstra, B. and J.W. Arntzen, *Description of a new species of crested newt, previously subsumed in Triturus ivanbureschi (Amphibia: Caudata: Salamandridae).* Zootaxa, 2016. **4109**(1): p. 73-80.
13. Wielstra, B., D. Salvi, and D. Canestrelli, *Genetic divergence across glacial refugia despite interglacial gene flow in a crested newt.* Evolutionary Biology, 2021. **48**(1): p. 17-26.
14. Rancilhac, L., et al., *Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders.* Molecular Phylogenetics and Evolution, 2021. **155**.
15. Horner, H.A. and H.C. Macgregor, *Normal development in newts Triturus and its arrest as a consequence of an unusual chromosomal situation.* Journal of Herpetology, 1985. **19**(2): p. 261-270.
16. De Visser, M.C., et al., *Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying Triturus' balanced lethal system* bioRxiv, 2024: p. 2024.10.25.620277.
17. De Visser, M.C., et al., *NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts).* bioRxiv, 2024: p. 2024.10.25.620290.
18. De Visser, M.C., et al., *PAV-spotter: using signal cross-correlations to identify Presence/Absence Variation in target capture data* bioRxiv, 2024: p. 2024.10.25.620064.
19. France, J., et al., *Genomic evidence suggests the balanced lethal system in Triturus newts originated in an instantaneous speciation event.* bioRxiv, 2024: p. 2024.10.29.620207.
20. Tamura, K., G. Stecher, and S. Kumar, *MEGA11: Molecular Evolutionary Genetics Analysis Version 11.* Molecular Biology and Evolution, 2021. **38**(7): p. 3022-3027.
21. Okonechnikov, K., et al., *Unipro UGENE: a unified bioinformatics toolkit.* Bioinformatics, 2012. **28**(8): p. 1166-7.
22. Untergasser, A., et al., *Primer3—new capabilities and interfaces.* Nucleic Acids Research, 2012. **40**(15): p. e115.
23. She, H., et al., *Identification of a male-specific region (MSR) in Spinacia oleracea.* Horticultural Plant Journal, 2021. **7**(4): p. 341-346.
24. Jiang, P., et al., *Cost-effective duplex Kompetitive Allele Specific PCR markers for homologous genes facilitating wheat breeding.* BMC Plant Biology, 2023. **23**(1): p. 119.
25. Suo, W., et al., *Towards low cost, multiplex clinical genotyping: 4-fluorescent Kompetitive Allele-Specific PCR and its application on pharmacogenetics.* PLoS ONE, 2020. **15**(3): p. e0230445.

26. Kalendar, R., et al., *Designing allele-specific competitive-extension PCR-based assays for high-throughput genotyping and gene characterization.* Frontiers in Molecular Biosciences, 2022. **9**: p. 773956.
27. Majeed, U., et al., *Kompetitive Allele Specific PCR (KASP): a singleplex genotyping platform and its application.* Journal of Agricultural Science, 2018. **11**(1).
28. Kaur, B., et al., *Utilization of KASP technology for wheat improvement.* Cereal Research Communications, 2020. **48**(4): p. 409-421.
29. Lear, G., et al., *Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples.* New Zealand Journal of Ecology, 2018.
30. Cavanaugh, S.E. and A.S. Bathrick, *Direct PCR amplification of forensic touch and other challenging DNA samples: A review.* Forensic Science International: Genetics, 2018. **32**: p. 40-49.
31. Brusa, A., et al., *A needle in a seedstack: an improved method for detection of rare alleles in bulk seed testing through KASP.* Pest Management Science, 2021. **77**(5): p. 2477-2484.
32. Hsiao, S.J., *Sources of error in molecular diagnostic analyses*, in *Accurate Results in the Clinical Laboratory*. 2019. p. 337-347.
33. Wielstra, B., et al., *Efficient screening for 'genetic pollution' in an anthropogenic crested newt hybrid zone.* Conservation Genetics Resources, 2016. **8**: p. 553-560.
34. Ertiro, B.T., et al., *Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize.* BMC Genomics, 2015. **16**: p. 908.
35. Krog, G.R., et al., *Blood group genotyping of blood donors: validation of a highly accurate routine method.* Transfusion, 2019. **59**(10): p. 3264-3274.
36. Altwayan, R., et al., *KASP-PCR method to screen thrombophilia genetic risk factors.* medRxiv, 2023.
37. Landoulsi, Z., et al., *Using KASP technique to screen LRRK2 G2019S mutation in a large Tunisian cohort.* BMC Medical Genetics, 2017. **18**(1): p. 70.
38. He, L., et al., *Simple, sensitive and robust chicken specific sexing assays, compliant with large scale analysis.* PLoS ONE, 2019. **14**(3): p. e0213033.
39. Bruijnis, M.R.N., et al., *Moral "lock-In" in responsible innovation: the ethical and social aspects of killing day-old chicks and its alternatives.* Journal of Agricultural and Environmental Ethics, 2015. **28**(5): p. 939-960.
40. Hattori, R.S., et al., *A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(8): p. 2955-9.
41. Geffroy, B. and C. Wedekind, *Effects of global warming on sex ratios in fishes.* Journal of Fish Biology, 2020. **97**(3): p. 596-606.
42. Ospina-Alvarez, N. and F. Piferrer, *Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change.* PLoS ONE, 2008. **3**(7): p. e2837.
43. Edmands, S., *Sex ratios in a warming world: thermal effects on sex-biased survival, sex determination, and sex reversal.* Journal of Heredity, 2021. **112**(2): p. 155-164.
44. Kupper, C., et al., *A supergene determines highly divergent male reproductive morphs in the ruff.* Nature Genetics, 2016. **48**(1): p. 79-83.
45. Tuttle, E.M., et al., *Divergence and functional degradation of a sex chromosome-like supergene.* Current Biology, 2016. **26**: p. 344-350.

Close-up Chioglossa lusitanica longipes*, Asturias, Spain*

Chapter 4 - NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts)

Manon Chantal de Visser[1,2], James France[1,2], Evan McCartney-Melstad[3,4,5], Gary Bucciarelli[4,5,6], Anagnostis Theodoropoulos[1,2], Howard Bradley Shaffer[4,5], Ben Wielstra[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands
2. Understanding Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands
3. Nutcracker Therapeutics, Emeryville, California, USA
4. Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, California
5. La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, UCLA, Los Angeles, California
6. Department of Wildlife, Fish, and Conservation Biology, University of California, Davis, California, USA

This manuscript has been submitted to the journal ***Molecular Ecology Resources*** and is currently under review. A preprinted version of this manuscript has been published at ***bioRxiv*** (2024.10.25.620290; https://doi.org/10.1101/2024.10.25.620290).

Abstract

Salamanders have large and complex genomes, hampering whole genome sequencing, but reduced representation sequencing provides a feasible alternative. We present NewtCap: a sequence capture bait set that targets c.7k coding regions across the genomes of all True Salamanders and Newts (the family Salamandridae, also known as 'salamandrids'). We test the efficacy of NewtCap, originally designed for the Eurasian *Triturus* newts, in 30 species, belonging to 17 different genera, that cover all main Salamandridae lineages. We also test NewtCap in two other salamander families. We discover that NewtCap performs well across all Salamandridae lineages (but not in the salamander families Ambystomatidae and Hynobiidae). As expected, the amount of genetic divergence from the genus *Triturus* correlates negatively to capture efficacy and mapping success. However, this does not impede our downstream analyses. We showcase the potential of NewtCap in the contexts of; 1) phylogenomics, by reconstructing the phylogeny of Salamandridae, 2) phylogeography, by sequencing the four closely related species comprising the genus *Taricha*, 3) hybrid zone analysis, by genotyping two *Lissotriton* species and different classes of interspecific hybrids, and 4) conservation genetics, by comparing *Triturus ivanbureschi* samples from several wild populations and one captive-bred population. Overall, NewtCap has the potential to boost straightforward, reproducible, and affordable genomic studies, tackling both fundamental and applied research questions across salamandrids.

Keywords

Hyb-seq, exon capture, Urodela, Caudata, High Throughput Sequencing, target enrichment

Introduction

One of the most challenging groups of animals to study genomically are the salamanders. These organisms have complex and large genomes that contain many repetitive elements compared to most other animals [e.g. they can be in the range of 10 to 40 times the size of a human genome; 1, 2-6]. This is the primary reason that conducting whole genome sequencing and *de novo* genome assembly for salamanders is extremely costly in terms of money, time, and computational resources [7, 8].

Fortunately, reduced representation sequencing strategies are paving the way toward more straightforward, reproducible, and affordable genomic studies, especially in organisms with large and complex genomes [9, 10]. By focusing sequencing efforts on subsets of the genome, rather than the entire genome, valuable time and resources can be conserved, allowing for a greater number of samples to be processed [11]. Different types of genome-subsampling techniques exist, broadly categorized as 'non-targeted' versus 'targeted', and their suitability varies, depending on the study species and research objectives [12-14].

Non-targeted approaches, such as Restriction site-Associated DNA sequencing (RAD-seq) and related techniques, are widely used in genetic mapping and population studies, including salamander studies [e.g. see; 15, 16-23]. While simple, scalable [11], and having the key advantage of not needing to know the sequences of any loci beforehand, such non-targeted approaches are known to yield missing data, underestimate genetic diversity, and call incorrect allele frequencies [24-26]. This is due to restriction site polymorphisms that limit the phylogenetic signal for resolving deep, evolutionary relationships [25, 27, 28].

On the other hand, targeted methods such as target capture sequencing offer a strategy to achieve higher resolution and specificity, as they focus on pre-selected (orthologous) loci [29, 30]. With target capture sequencing, biotinylated RNA probes – or 'baits' – that are complementary to the loci of interest are used, causing them to bind to the 'target' regions, before streptavidin-coated magnetic beads are used to 'capture' them [31, 32]. The main advantage of using the target capture method over untargeted methods is that there will be higher efficiency, because enrichment of specific (usually coding) genomic regions of interest is more effectively achieved across samples [11, 33-35]. Furthermore, the flanking regions of targets can also provide information on more variable genomic regions such as introns [14, 36], and off-target 'bycatch' reads can provide additional data as well [37, 38].

Often custom target capture baits are designed, which can be based on a draft genome or transcriptome reference, however pre-designed baits can be ordered as well [11, 39, 40]. Over the last decade, many target capture bait sets have been tested and

made publicly available for different types of organisms, ranging from micro-organisms to macro-organisms [e.g.; 11, 33, 34, 41-43]. For animals in particular, bait sets exist for certain groups of insects [e.g. Hymenoptera; 44], snails [e.g. Eupulmonata; 45], reptiles [e.g. Squamata; 46, 47], fish [e.g. Acanthomorpha; 48], and amphibians [e.g. Anura; 49]. Salamander bait sets have also already been designed for particular genera [Ambystoma and *Triturus*; 50, 51]. As the target capture approach generally handles a certain degree of sequence divergence well, a bait set designed for one genus has great potential to work in other genera too [52, 53]. Thus, it is worth exploring the potential transferability of such existing bait sets to related taxa.

We introduce 'NewtCap': a target capture bait set and protocol for salamanders of the family Salamandridae (which includes the 'True Salamanders' and the 'Newts'). The bait set, originally designed for *Triturus* newts [51], targets c. 7k putative, orthologous, coding regions. Here, we provide an updated version of the lab protocol that cuts down on costs, time and DNA input. Furthermore, we investigate the efficacy of NewtCap across 30 different species and 17 genera of the Salamandridae family and test its efficiency in two distantly related species of other salamander families. Besides checking general performance, we assess the usefulness of NewtCap in the contexts of; 1) phylogenomics, by inferring a Salamandridae phylogeny, 2) phylogeography, by investigating the relationships among multiple individuals of four closely related *Taricha* species, 3) hybrid zone analysis, by calculating the hybrid index and heterozygosity for two *Lissotriton* species and interspecific hybrids of different cross types, and 4) conservation genetics, by determining the genetic relatedness of a captive-bred *Triturus ivanbureschi* population to wild populations. Overall, we demonstrate that NewtCap is an important tool for molecular studies on salamandrids.

Materials and Methods

Sampling & DNA extraction

We studied 73 individual salamanders in total, covering 30 species and 17 genera of the Salamandridae family (including nine interspecific hybrids of the genus *Lissotriton*), as well as two more distantly related samples (Table S1, see Zenodo) from the families Ambystomatidae and Hynobiidae [54, 55]. We obtained DNA extractions or genetic data from previous studies [51, 56-64], as well as new samples for this study (provided by collaborators, see acknowledgments and Table S1 via Zenodo).

We used the Promega Wizard™ Genomic DNA Purification Kit (Promega, Madison, WI, USA), which is a salt-based DNA extraction protocol [65]. The DNA from each sample was re-suspended in 100 µL 1 × TE buffer before concentrations and purity were assessed via spectrophotometry, using the DropSense96™ (Trinean, Gentbrugge,

Belgium). As we used a minimum concentration of 150 ng/µL for library preparation, any sample found to be below was concentrated by vacuum centrifugation.

NewtCap: Laboratory procedures

The target capture procedures are fully documented and included as Supplementary Materials (see under 'Data Accessibility'). These protocols are rigorously optimized versions of those previously described [see; 51]. Sonication of genomic DNA has been replaced with enzymatic fragmentation, resulting in a tenfold increase in library yield per ng of input DNA, in addition to time and cost savings. The volumes of all reagents in the library preparation were reduced by 75% compared to the manufacturers protocol to conserve reagents. Target capture protocols have been adapted from Mybaits V3.0 to the V4.0 kit and are fully compatible with Mybaits V5.0.

Library preparation

DNA libraries were constructed using the NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) following the manufacturers protocol, with quarter volumes of all reagents and with the following modifications: 1,000 ng of extracted genomic DNA was used as input (6.5 µL at 154 ng/µL). The enzymatic shearing time at 37 °C was adjusted to 6.5 minutes (as the minimum time of 15 minutes suggested by the manufacturer resulted in over-digestion). After NEB adapter ligation and cleavage with the NEB USER enzyme, NucleoMag™ magnetic separation beads (Macherey-Nagel, Düren, Germany) were used for double-ended size selection targeting an insert size of 300 bp. Libraries were indexed with 8 cycles of PCR amplification, using unique combinations of custom i5 and i7 index primers (Integrated DNA Technologies, Leuven, Belgium). NucleoMag™ beads were used again for a final cleanup before the libraries were resuspended in 22 µL of 0.1 × TE buffer. Library size distribution and concentration was measured using the Agilent 4150 TapeStation or 5200 Fragment analyzer system (Agilent Technologies, Santa Clara, CA, USA), using the D5000 ScreenTape or DNF-910 dsDNA Reagent Kit. We aimed for obtaining a final library concentration of at least 12 ng/µL.

Target capture, enrichment, & sequencing

Libraries were equimolarly pooled in batches of 16, aiming for a total DNA mass of 4,000 ng (250 ng per sample). Vacuum centrifugation was then used to reduce the volume of each pool to 7.2 µL (556 ng/µL). We performed target capture with the MyBaits v4.0 kit (Arbor Biosciences, Ann Arbor, MI, USA) previously designed for *Triturus* newts (Wielstra et al., 2019), which targets 7,139 unique exonic regions (product Ref: # 170210-32). The manufacturers protocol was employed with the following deviations: Blocks C and O were replaced with 5 µL of *Triturus* derived C0t-1 DNA at 6,000 ng/µL (30,000 ng per pool). C0t-

1 DNA is enriched in repetitive sequences and acts as a blocking buffer to non-specific targets in capture assays by hybridizing with repetitive sequences in the libraries (McCartney-Melstad et al., 2016). Tissue to produce C0t-1 DNA was derived from an invasive population of *T. carnifex* (Meilink et al., 2015).

The pooled libraries were incubated with the blocking buffer for 30 minutes, followed by hybridization for 30 hours at 62 °C. After capture of the hybridized baits with streptavidin coated beads and four cycles of washing, each pool was divided into equal-volume halves. The first half was subject to 14 cycles of PCR amplification, followed by bead cleanup, and resuspension in 22 µL of 0.1 × TE buffer. The concentration and fragment size distribution of the enriched pool were then measured with the TapeStation system, using the HS D5000 ScreenTape kit. If the final concentration was between 5 and 20 nM then the same protocol was employed for the second half of the pool. If not, the number of post-enrichment PCR cycles was altered to compensate. For each pool 16 GB (1 GB per sample) of 150 bp paired-end sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) by BaseClear B.V. (Leiden, the Netherlands).

Bioinformatics for data pre-processing

We followed a standard pipeline for cleaning up the raw reads and mapping the trimmed sequence data against reference sequences in a Linux environment [adapted from; 51]. We describe the main steps and provide the scripts we used publicly through GitHub: https://github.com/Wielstra-Lab/NewtCap_bioinformatics.

Quality control & Read clean-up

First, we clipped all reads to a maximum length of 150 bp using the BBDuk script by BBTools (BBMap – Bushnell B. – https://sourceforge.net/projects/bbmap/). Then we removed any leftover adapter contamination and low-quality bases (or reads) using Trimmomatic v.0.38 [66]. Adapter sequences for the TruSeq2 multiplexed libraries were identified and removed. Leading and trailing bases were trimmed if the Phred score was <5. We also removed reads in case the average Phred score in a sliding window (5' to 3') of a size of five base pairs dropped below 20, and we discarded all reads shorter than 50bp. We monitored the quality of our sequences before and after trimming with FastQC [67] by using the summarizing *quality_check* function of SECAPR [68].

Read mapping & Variant calling

Cleaned reads were mapped to the set of 7,139 *T. dobrogicus* reference sequences with a maximum length of 450bp [based on T. dobrogicus transcripts, see; 51] that were initially used for probe development. The reference FASTA file is provided as Supplementary

Material. Mapping was performed using the MEM algorithm from Burrows-Wheeler Aligner v.0.7.17 [69], and we stored results in BAM format using SAMtools v.1.18 [70, 71]. We added read group information using the *AddOrReplaceReadGroups* function of Picard v.2.25.1 and PCR duplicates were flagged with the Picard function *MarkDuplicates* (http://broadinstitute.github.io/picard/). Afterwards, the output was again saved in BAM format. Next, we called variants using the *HaplotypeCaller* function of GATK v.4.1.3.0 [72] including the *-ERC GVCF* option, and we used GATK's *CombineGVCFs* and *GenotypeGVCFs* functions to perform joint genotype calling to create multi-sample (ms) gVCF files as input for downstream analyses. In case samples still needed to be added or removed from msgVCF files after joint genotyping, we did so using the *view* function of BCFtools v.1.15.1 [71, 73].

Downstream analyses

Bait performance statistics

To evaluate the overall performance of the NewtCap baits, we calculated several statistics (Table S1). First, we counted the number of GB and reads contained in the raw FASTQ files for each sample. We also used the SAMtools' *flagstat* function to determine the total number of reads present in the (deduplicated) BAM files that passed quality control, as well as the percentage of these reads that successfully mapped to a reference sequence. We used SAMtools' *coverage* function to extract information on the mean depth of coverage, as well as the mean percentage of coverage, for each target, and then averaged these per sample to provide an estimate of the "success rate". Besides analyzing the contents of the BAM files, we checked how many sites were outputted in total in each separate gVCF file by counting lines (as each non-header line is one site). Then, we counted how many of those were considered SNPs, and how many were considered INDELs, by using the *stats* function of BCFtools (note that this is done before merging any files or applying any SNP filtering).

To estimate how the performance of NewtCap correlates with the level of genetic divergence from *T. dobrogicus*, the species that was used for bait design and as a reference for read mapping, we performed several statistical analyses. For these, we used a rough estimate of divergence times for lineages of Salamandridae [from; 56]. We explored the relationship between this estimated genetic divergence from *T. dobrogicus* and the following performance variables; 1) the amount of GB of data obtained, 2) the number of obtained reads, 3) the percentage of mapped reads, 4) the percentage of reads marked as PCR duplicates, 5) the mean read depth number after deduplication, 6) the mean coverage of the sequence bases after deduplication, 7) the number of SNPs found in the gVCFs files, and 8) the number of INDELs found in the gVCF files. We calculated the correlation coefficients (r) between either of these variables and the estimated genetic

divergence from *T. dobrogicus*. Depending on whether our data met the assumptions for parametric or non-parametric testing, we either employed the Pearson correlation method or the Spearman's rank correlation method. We determined the level of significance using a two-tailed test, with the p-value threshold of $p<0.00625$ to indicate statistical significance (with a Bonferroni correction on the usual threshold $p<0.05$ for the eight tests performed, as 0.05/8 = 0.00625). These analyses, including assumption checks such as testing for normality, were done in Microsoft Excel 2024.

Concatenated phylogenetic analyses in RAxML

To investigate the usefulness of NewtCap in the context of phylogenomics, we built phylogenetic trees with a Maximum Likelihood method using RAxML. We used NewtCap to reconstruct an existing Salamandridae tree that was originally built based on 5,455 nuclear genes derived from transcriptome data [74], by including 23 samples that cover at least one representative genus for each of the main clades *in sensu* [as described in; 74]. To further check the performance of NewtCap across True Salamanders as well as for non-salamandrid species, and to explore the position of the root, we added three additional samples in a second, extended analysis, namely the Salamandridae species *Mertensiella caucasica* (family Salamandridae) and two non-salamandrid species *Ambystoma mexicanum* (family Ambystomatidae) and *Paradactylodon gorganensis* (family Hynobiidae). The raw msgVCF files for these two analyses respectively comprised 812,574 sites and 812,603 sites from across 7,135 targets. Note that not all sites in the raw and intermediate (ms)gVCF files were necessarily polymorphic (i.e. invariant - or monomorphic - sites may still have been included, unless we specifically stated that these have been removed).

We applied quality filtering on the msgVCFs (which contained only the samples chosen for these phylogenetic analyses). First, we removed sites that showed heterozygote excess ($p<0.05$) using BCFtools v.1.15.1 , in order to rid paralogous targets [see; 51]. Then, we used VCFtools v.0.1.16 to enforce the following strict filtering options; discarding INDELs ("--remove-indels"), filtering out sites with for instance poor genotype and mapping quality scores ["QD<2", "MQ<40", "FS>60", "MQRankSum <−12.5", "ReadPosRankSum<−8", and "QUAL<30"; 75], and discarding sites with over 50% missing data ("--max-missing = 0.5"). At this stage, the intermediate msgVCF files contained, in total, 625,182 SNPs from across 7,103 targets, and 621,754 SNPs from across 7,095 targets, for the sets of 23 and 26 samples.

We converted the files into PHYLIP format with the 'vcf2phylip.py' script [76], which, by default, requires a minimum of four samples representing each SNP variant. Next, we performed an ascertainment bias correction by using the 'ascbias.py' script (https://github.com/btmartin721/raxml_ascbias) to remove sites considered invariable by RAxML. The final PHYLIP file was used as input for RAxML v.8.2.12 [77], which we ran

with 100 rapid bootstrap replicates, under the ASC_GTRGAMMA model, and with the Lewis ascertainment correction [78]. We obtained the best-scoring Maximum Likelihood tree out of the concatenation analyses.

To gain insights into the potential of NewtCap in a phylogeographical context, we also built a phylogenetic tree for the New World newt genus *Taricha* using RAxML. As input, we used a msgVCF file containing nine samples: eight *Taricha* samples covering four distinct species within the genus (*T. granulosa*, *T. rivularis*, *T. sierrae* and *T. torosa*, with two samples per species) and one sample of the sister-genus *Notophthalmus* as outgroup (Table S1). The raw msgVCF file comprised 209,072 sites from 7,059 targets. Subsequently, we performed the filtering steps as described above, which left the intermediate msgVCF file with 180,385 SNPs from across 7,048 targets. We visualized all phylogenies using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Hybrid analyses & Conservation genetics

To assess the usefulness of NewtCap in the context of hybridization studies, we used the R packages '*triangulaR*' [79], '*ggplot2*' [80] and '*vcfR*' [81] to build triangle plots. We investigated the target capture data of in total 15 *Lissotriton* individuals; three of each parental species (*L. vulgaris* and *L. montandoni*), as well as hybrids bred and reared in the lab: three F1 hybrids, three F2 hybrids, and three backcrosses (F1 x *L. vulgaris*). We assembled a raw msgVCF file containing only these samples, which initially had data on 807,389 sites across in total 7,135 targets. From this we extracted high quality SNPs as described before (except here we tolerated no missing data at all). The filtered msgVCF contained 532,332 SNPs from 6,866 targets in total and was used as input for *triangulaR*.

The *triangulaR* package uses SNPs from an input VCF file that are estimated to be species-diagnostic based on the parent species under a certain allele frequency threshold between '0' and '1', where '1' equals a fixed difference between parental population [79]. Thus, the threshold of 1 is employed when searching for two consistent, diverged, homozygous states in the parental populations. SNPs that pass this filter can then be used to calculate the hybrid index and heterozygosity for the samples analyzed. Considering that our parent populations did not comprise the actual parents of the hybrids included in our study and were represented by just three samples each, this complicated distinguishing species-diagnostic SNPs from SNPs that only appeared diagnostic by chance. Therefore, we only extracted SNPs that were 100% species-diagnostic (based on the information from the parental populations, by setting the allele frequency threshold to 1) *and* that were always heterozygous in the F1's. This extra functionality (i.e. filtering for heterozygosity in F1 hybrids) is currently not built into *triangulaR*, but we added this filtering option before conducting our calculations. We provide the customized R script in our GitHub repository.

To assess the performance of NewtCap in the context of a conservation genetic study we conducted a Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) to determine the geographical origin of a captive bred population, which is presumed to originate from Cerkezköy in Turkey (Michael Fahrbach, pers. comm.). We used the R packages '*gdsfmt*' and '*SNPRelate*' [82] on data from 24 *Triturus ivanbureschi* newts that originated from seven different wild populations and the captive population (i.e. three samples per population, see Table S1). Populations from the wild include both the glacial refugial area (Asia Minor and Turkish Thrace) and postglacially colonized area [the Balkans; 83]. We created a raw msgVCF file from these samples, which contained a total of 812,768 sites from across 7,135 targets. We extracted high quality SNPs in the same way as we did for the hybrid studies, and the filtered msgVCF that was used as input had a total of 486,891 SNPs from 5,774 targets. We conducted the hybrid and conservation genetic analyses in Rstudio [84] using R v.4.1.2 [85].

Finally, we were interested in the number of informative SNPs that we could find within Salamandridae species that are distantly related to *T. dobrogicus*. The most distantly related species for which we had more than one sample belonging to distinct populations in our dataset is *Chioglossa lusitanica* (Table S1), which is a true salamander, not a newt. Having more than one sample allows us to cross-compare these populations. Thus, we extracted the variant information of the two *C. lusitanica* samples from the overall msgVCF file –filtered for high quality SNPs allowing no missing genotypes– that was used to estimate the genetic divergence from *T. dobrogicus* [56]. We quantified the number of SNPs that were homozygous in one sample and heterozygous in the other sample – and *vice versa*. Additionally, we counted the number of SNPs that were homozygous in both samples, but for different alleles.

Results

Overall performance of NewtCap

In total we collected 149.0 GB of raw sequence data with on average 15,573,781 reads per sample (s.d. = 11,400,957). On average 42.6% (s.d. = 16.2%) of all reads were successfully mapped against the *Triturus* reference, and an average of 11.9% (s.d. = 10.6%) of all reads were marked as PCR duplicates. In the deduplicated BAM files, the mean read depth across all targets per sample varied widely, and ranged from 5.9 X (s.d. = 27.9) in *A. mexicanum* and 8.2 X (s.d. = 27.8) in *P. gorganensis* to 122.2 X (s.d. = 345) in one of the *Lissotriton* hybrids. The mean coverage of the sequence bases of all targets

ranged from 36.1% (s.d. = 30.3%) in *P. gorganensis* and 36.4% (s.d. = 31.3%) in *A. mexicanum* to 98.31% in two *T. ivanbureschi* samples (s.d. = 7.2% and s.d. = 6.9%). On average, the mean read depth was 60.5 X (s.d. = 183.2) across all samples, and the overall mean coverage of the sequence bases was 92.0% (s.d. = 13.7%). Details are in Table S1.

We discovered significant correlations between the genetic divergence from *T. dobrogicus* and several performance variables. Firstly, as the divergence from *T. dobrogicus* increases, so does the amount of sequence data gathered, both as measured in GB in the raw FASTQ files and as measured in the total number of 'QC-passed' reads observed by SAMtools (Spearman's rank correlation test, $p<0.001$ for both, Table 1). We found no significant correlation between the percentage of reads marked as PCR duplicates by Picard and the level of divergence from *T. dobrogicus* ($p = 0.91$), however we did find a negative correlation between the percentage of reads that map against the *T. dobrogicus* reference sequences and the amount of divergence from *T. dobrogicus* (Spearman's rank correlation test, $p<0.001$, Table 1). Furthermore, there is a negative correlation between the coverage of reference targets and the level of divergence from *T. dobrogicus* - i.e. the mean read depth and the mean coverage of the sequence bases both drop as the divergence from *T. dobrogicus* increases (Pearson correlation test for the mean read depth, $p<0.001$ and Spearman's rank correlation test for the mean coverage of the sequence bases, $p<0.001$; Table 1). Lastly, with increased divergence from *T. dobrogicus*, more SNPs and INDELs are discovered by BCFtools in the eventual gVCF files (Spearman's rank correlation test, $p<0.001$ for both, Table 1).

Phylogenomics: Reconstruction of the Salamandridae phylogeny

The fully resolved and highly supported Salamandridae phylogeny resulting from Maximum Likelihood analysis of concatenated data in RAxML was based on 204,600 polymorphic SNPs (Fig. 1 and Fig. S1). The placement of the root on the branch connecting the newts and the clade containing the True Salamanders and *Salamandrina* was confirmed by our extended phylogeny (based on 265,105 SNPs, Fig. S2) that included two non-Salamandridae salamander species (*A. mexicanum* and *P. gorganensis*), as well as a True Salamander species (*M. caucasica*).

Phylogeography: Fully resolved relationships of *Taricha*

The *Taricha* phylogeny that resulted from the concatenated data used by RAxML to perform a Maximum Likelihood analysis, was based on 9,730 polymorphic SNPs. The tree was fully resolved and highly supported (Fig. 2).

***Table 1: Correlation statistics for the relationship between NewtCap performance variables and the estimated amount of species divergence from* Triturus dobrogicus.** *For each of the separate performance variable the appropriate statistical test - either the Spearmann's rank or the Pearson correlation test - was applied to determine whether the observed correlations were significant. Both the correlation coefficient (R) and its square value are provided, as well as the p-value. Results with a p-value lower than 0.00625 (including a Bonferroni correction, see Methods) are marked with an asterisk.*

Performance Variable	Test applied	R	R^2	P-value
Amount of raw data R1 + R2 (in GB)	Spearmann	0.610	0.372	1.6E-08*
Number of QC-passed reads	Spearmann	0.641	0.411	1.7E-09*
Percentage of reads mapped	Spearmann	-0.661	0.440	3.6E-10*
Percentage of PCR duplicates	Spearmann	0.015	<0.001	0.910
Mean read depth	Pearson	-0.597	0.357	3.8E-08*
Mean coverage of sequence bases	Spearmann	-0.918	0.842	2.2E-29*
Number of SNPs discovered	Spearmann	0.917	0.841	2.7E-29*
Number of INDELs discovered	Spearmann	0.899	0.807	2.4E-26*

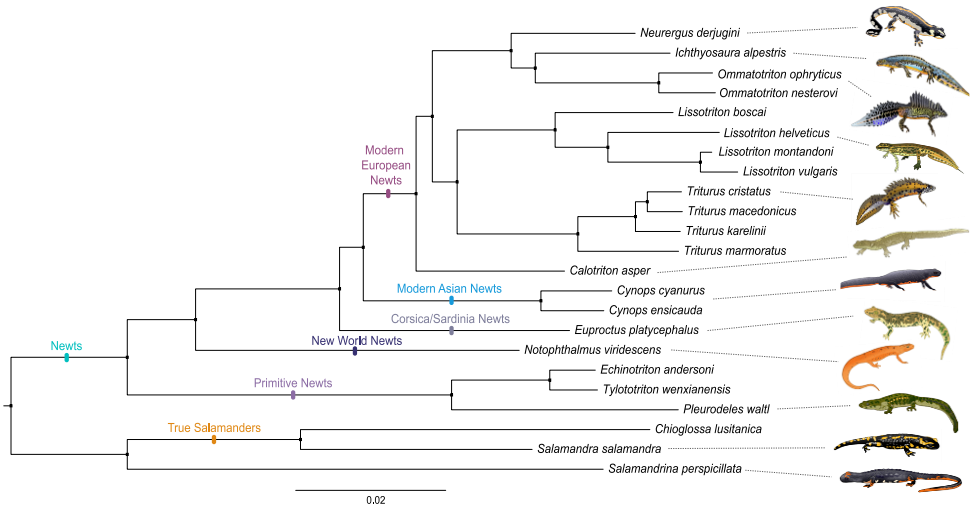


Figure 1: NewtCap-based phylogeny of the Salamandridae family. *The phylogeny is based on Maximum Likelihood inference of concatenated data of 204,600 informative SNPs using RAxML. Overall layout and clade labels conform to a previous transcriptome-based phylogeny [74]. The tree is rooted on the branch separating the newts and the clade containing the True Salamanders and* Salamandrina *(see also Fig. S1 for the same tree, but with original labels, and Fig. S2 for the additional, extended tree, including* Ambstoma, Paradactylodon *and* Mertensiella, *that confirms the root position adopted here). All nodes have a bootstrap support of 100%.*

Hybridization studies: *Lissotriton* hybrids & backcrosses detected

After quality filtering, we identified 666 SNPs that we consider species-diagnostic (i.e. 'ancestry-informative') for *L. vulgaris* versus *L. montandoni* in the target capture data based on the genotypes of the six parental species samples and three F1 hybrids. Those SNPs enabled us to calculate the hybrid indices and interclass heterozygosity values in F2 and backcross ('Bx') hybrids, as visualized in a triangle plot (Fig. 3).

Conservation genetics: Separation of *Triturus* populations & *Chioglossa* subspecies

For the PCA and HCA analyses, the R calculations ended up being based on 9,135 bi-allelic SNPs. Along the first and second Principal Components of the PCA, the four wild populations from the postglacially colonized area cluster together, whereas the three wild populations from the glacial refugial area stand relatively apart Fig. 4A & B). The captive-bred individuals cluster closest to the Safaalan group, a population in Turkey just west of Istanbul, close to the presumed source locality Cerkezköy. The same pattern is observed in the HCA dendogram (Fig. 4C).

The check for high quality SNPs that display different genotypes in the two *C. lusitanica* samples originating from different subspecies, resulted in a list of over 10,000 polymorphic SNPs: we counted 8,029 SNPs for which one individual was homozygous, and the other individual was heterozygous, and we discovered an additional 2,301 informative SNPs where the individuals were both homozygous, but for alternate alleles.

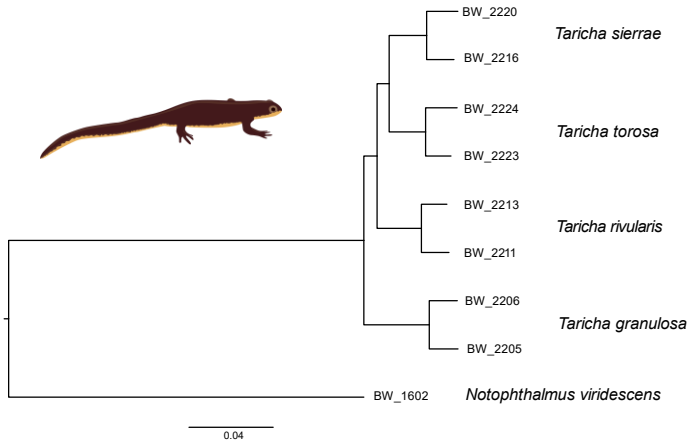


Figure 2: A* Taricha *phylogeny obtained with NewtCap-derived data. *The phylogeny is based on Maximum Likelihood inference of concatenated data of 9,730 informative SNPs using RAxML.* Notophthalmus *is used to root the tree. All nodes have a bootstrap support of 100% (not shown).*

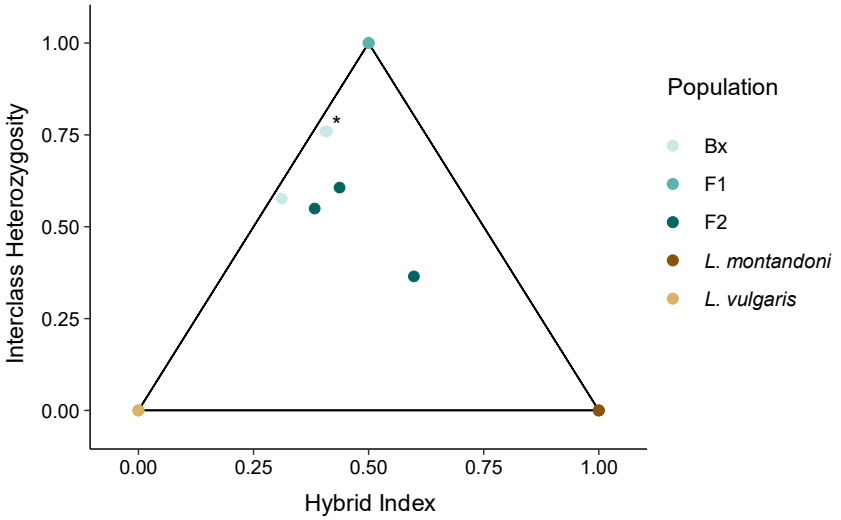


Figure 3: Triangle plot of different* Lissotriton *hybrid classes based on NewtCap-derived data. *The plot, based on 666 informative SNPs, shows the relationship between the hybrid index (the fraction of the alleles per individual that derived from each of the two parental species, also known as the ancestry) and the interclass heterozygosity (the fraction of the alleles per individual that is heterozygous for alleles from both parental species). The* L. vulgaris *individuals are in the bottom left corner, the* L. montandoni *individuals in the bottom right corner, and the F1 hybrid offspring in the top corner. The F2 and Bx ('backcross') hybrids are placed inside the triangle, with two Bx samples almost fully overlapping (marked with *).*

Discussion

We introduce NewtCap, a target capture bait and reference set of 7,139 sequences applicable to salamandrids. We show that NewtCap works effectively across all main lineages within the Salamandridae family. As anticipated, the target capture rates and mapping successes are influenced by the level of genetic divergence from *T. dobrogicus* and we see that – within Salamandridae – more off-target regions are getting captured for more distantly related species. However, these influences appear to be minor and evidently do not hamper our downstream analyses: only for salamanders outside of the family Salamandridae does NewtCap provide data of insufficient quality. NewtCap thus proves to be a powerful tool to collect genomic data for Salamandridae studies regarding systematics (e.g. phylogenomics, phylogeography) and population genetics (e.g. hybrid studies, conservation genetics).

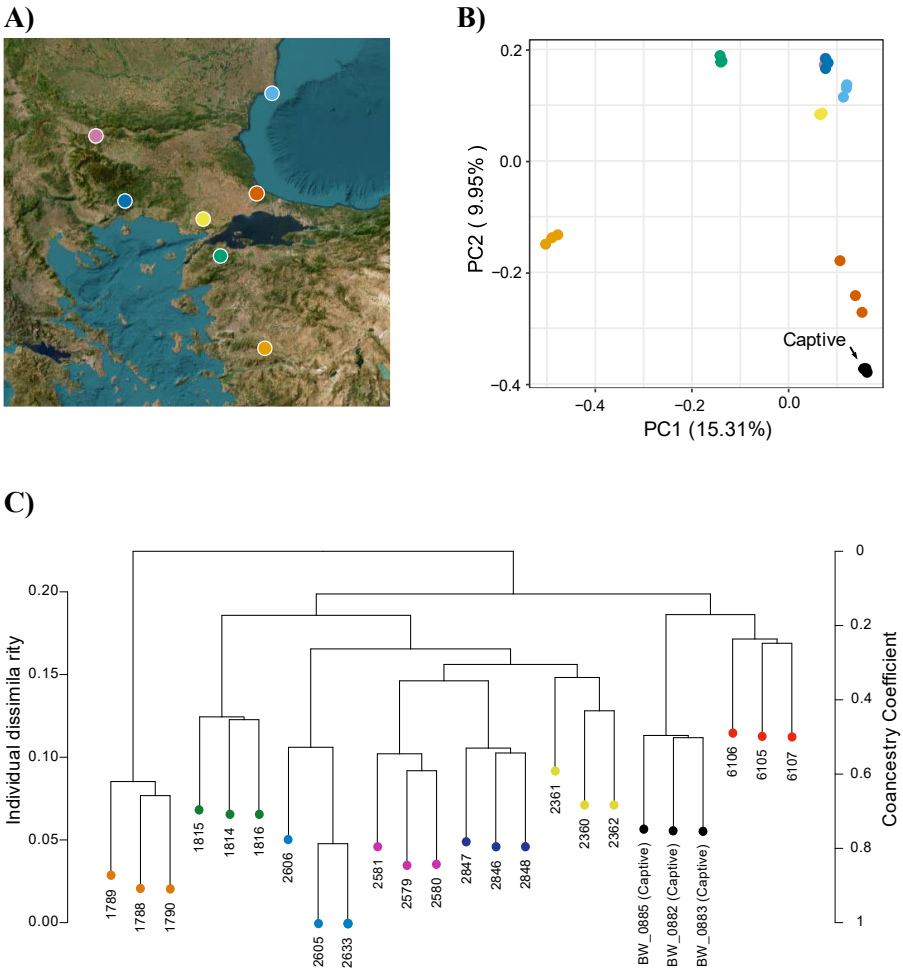


Figure 4: Genetic differentiation between wild and captive* Triturus ivanbureschi *populations based on NewtCap-derived data. (A) *The wild population localities (details in Table S1; four postglacial populations are represented by dark blue, light blue, yellow, and pink colors in Bulgaria, Greece and Turkey, and three populations from the glacial refugial area are represented by red, orange and green colors in Turkey).* ***(B)*** *A plot of the first versus the second Principal Component (PC) places the captive individuals closest to a population from just west of Istanbul.* ***(C)*** *The dendrogram produced by the HCA analysis, showing the Individual Dissimilarity as well as the Coancestry Coefficient, again shows that captive samples cluster with a population just west of Istanbul.*

We encourage potential users to further tweak the NewtCap workflow. For instance, in terms of the laboratory protocol, researchers could consider; 1) replacing the C0t-1 blocker with a standard blocker, 2) reducing the hybridization time, and 3) including more individuals per capture reaction. In terms of bioinformatics, the lower mapping rate observed in Salamandridae species more distantly related to *T. dobrogicus* does not necessarily only reflect decreased enrichment efficiency due to genetic divergence; mapping rate is presumably also influenced by the reduced ability of the read mapper to align sequencing reads to more divergent reference sequences used in the bioinformatics pipeline [52, 68]. Users could explore applying different mapping settings, using alternative mapper tools, or making use of (or generating) substitute reference sequences for read alignment [11, 86].

Our findings demonstrate that the current NewtCap protocol can effectively be applied to any member of the family Salamandridae – the crown of which is dated as far back as c. 100 MYA [56, 87]. Our Salamandridae phylogeny perfectly matches the topology of a transcriptome-based phylogeny (Rancilhac et al., 2021). Although a higher number of molecular markers does not necessarily result in a more accurate species tree [88, 89], independent RAxML analyses using different subsets of NewtCap data result in the same topology [with one notable exception, see; 56, 64].

NewtCap has already recently been applied to study the systematics and taxonomy of certain modern European newts – including the genera; *Triturus* [51, 57], *Lissotriton*, [58], and *Ommatotriton*, [59]. However, we show that NewtCap allows for a (putative) genome-wide subsampling of thousands of markers for other Salamandridae lineages as well. Existing phylogenies of modern Asian newts (genera *Cynops*, *Paramesotriton*, and *Pachytriton*; Fig. 1) rely on a limited amount of molecular markers and frequently fail to recover genera within the clade as monophyletic groups, presumably due to the intricate biogeographical history of this clade [90-92]. The New World Newts (genera *Taricha* and *Notophthalmus*; Fig. 1) have so far only been studied based on mtDNA and allozyme data [93-97]. Our *Taricha* phylogeny, the first one based on a considerable number of nuclear markers/ SNPs, comprises a fully supported positioning of each of the four known species, with a topology that matches that of an existing mtDNA-based phylogeny [98, 99]. For the Asian members of the primitive newts (genera *Tylototriton* and *Echinotriton*; see Fig. 1), genetic resources have been scarce so far, which is why there is an outstanding call to conduct more extensive, genomic research in order to better understand the evolution and taxonomy of the (sub)species of these lineages [100, 101].

NewtCap performs well even for the sister clade of the newts, containing the 'True Salamanders' (*Salamandra, Chioglossa, Mertensiella, Lyciasalamandra*; see Fig. 1) and the spectacled salamanders (*Salamandrina*). First, NewtCap provides empirical support for the recent suggestion that *Salamandrina* represents the sister lineage to the True Salamanders instead of to all remaining salamandrids [74]. *Salamandrina* itself has

so far only been studied with a relatively small number of markers [91, 101-106]. *Chioglossa* is a monotypic genus that is close to becoming threatened according to the IUCN Red List [107]. Conservationists generally identify monotypic taxa as 'evolutionarily unique', which helps justify elevated conservation imperatives [108-110]. We obtained c. ten thousand informative SNPs distinguishing two individuals belonging to different *C. lusitanica* subspecies [63]. The sister genus of *Chioglossa*, *Mertensiella*, is monotypic as well, but so far few populations have been studied and only with mtDNA [111, 112].

Next to accentuating the potential of NewtCap in the context of conservation genetics, we emphasize that the tool can also be used to identify the degree of interspecific gene flow and geographical structuring. For example, to test hypotheses about species status and historical biogeography. Introgressive hybridization is increasingly recognized as a source of adaptive variation in natural populations, [113-115], but also as a source of genetic pollution in the case where non-native and native individuals interbreed in nature [116-119]. We identify different hybrid classes of *Lissotriton* newts and are able to differentiate between genetically distinct groups of *T. ivanbureschi* – and determined to which wild population a known, captive population is genetically most similar. Such applications are valuable for guiding wildlife management practices and conservation efforts that concern Salamandridae species both *in situ* and *ex situ* [120-122]. This is especially important considering that, out of all vertebrate groups, amphibians are facing the most drastic population declines and extinction rates observed in the Anthropocene [123-127]. Overall, by providing in the range of thousands to hundreds of thousands of high-quality SNPs, NewtCap facilitates the molecular study of salamandrids whilst whole genome sequencing of gigantic salamander genomes remains unattainable.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the European Research Council grant agreement No. 802759 and the Marie Skłodowska-Curie grant agreement No. 655487. This work was performed using the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. Daniele Canestrelli, Andrea Chiocchio, Michael Fahrbach, Willem Meilink, and Frank Pasmans contributed samples. The tissue sample of *Notophthalmus viridescens* (SMF 99040) was donated by the Senckenberg Research Institute and Nature Museum. The *Taricha* tissue samples were collected following protocols for animal welfare approved by UCLA and permitted by California Department of Fish and Wildlife (SCP # 12430). We would like to thank Erik-Jan Bosch for making the scientific drawings of the salamanders used in Figure 1 under the

open content License of Naturalis Biodiversity Center (© CC BY-NC-ND 4.0), Benjamin Wiens for providing support with *triangulaR* adjustments, and Dr. Peter Scott for providing input on bioinformatic analyses. The *Triturus ivanbureschi* geographical map cut-out (Figure 4A) was made thanks to OpenStreetMap and its contributors (® CC BY-SA 2.0), and we obtained the *Taricha* cartoon from Figure 2 copyright free via user 'clker-free-vector-images-3736' through Pixabay.com.

Data accessibility & Benefit-sharing statement

The samples used in this study are in compliance with national laws and the Nagoya Protocol and detailed information can be found in the Supplementary Materials, via our Zenodo repository (which includes laboratory protocols, files required for bioinformatics, and the FASTA file containing the sequences that were used for initial probe tiling; https://doi.org/10.5281/zenodo.13785684). The main steps and scripts for the bioinformatics are provided on our GitHub repository (https://github.com/Wielstra-Lab/NewtCap_bioinformatics). Raw sequencing reads are accessible via BioProject 'PRJNA1171613' (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1171613).

Author contributions

B.W., M.d.V., & J.F. conceived and designed the research. B.W., E.M., & H.B.S. initially designed the NewtCap probes for *Triturus*, and J.F. & B.W. further optimized the tool for broader use and higher efficiency. M.d.V., J.F., & A.T. performed the lab-work. M.d.V. conducted the pre-processing of the data, as well as all downstream analyses. G.B. contributed to the analyses of *Taricha*. M.d.V., B.W., & J.F. wrote the draft version of the manuscript, and all authors contributed to revising it.

Conflict of interest

All authors declare no conflict of interest.

Supplementary figures

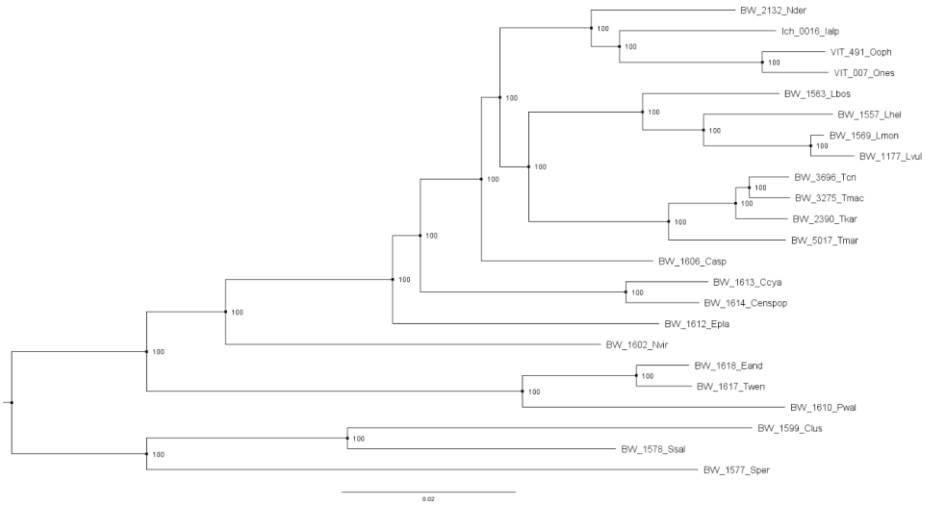


Figure S1: The raw, reconstructed NewtCap-based phylogeny of the Salamandridae family. *This is the same tree as provided in MS Fig. 1, but with original sample identifiers and bootstrap values. The tree is rooted on the branch separating the newts and the clade containing the True Salamanders and* Salamandrina *(see also Fig. S2, which confirms the root position adopted here).*

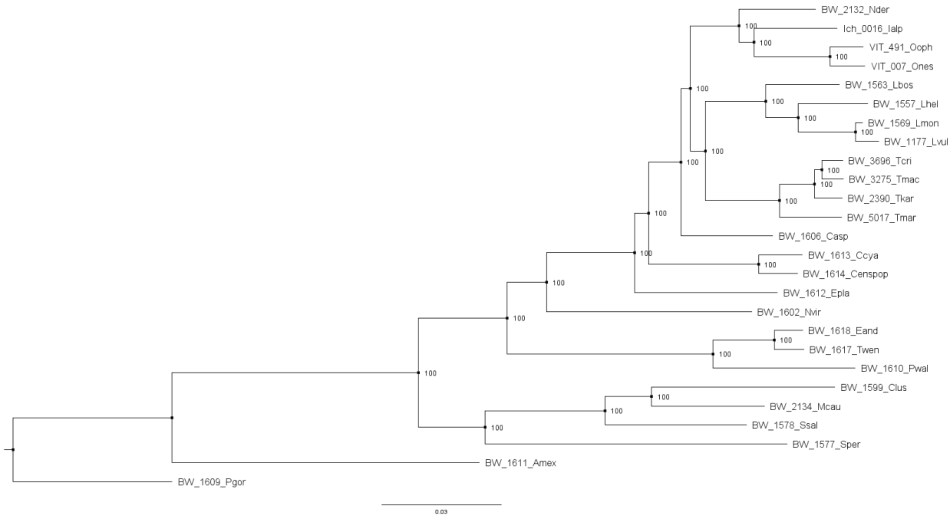


Figure S2: The raw, reconstructed NewtCap-based phylogeny of the Salamandridae family, including two non-salamandrids. *This tree is the result of an independent RAxML analysis that also includes a* Mertensiella caucasica *individual, and two non-salamandrid individuals: one* Ambystoma mexicanum *sample and one* Paradactylodon gorganensis *sample. The tree is based on 265,105 SNPs and shows sample identifiers and bootstrap values. The tree is rooted on the branch of* P. gorganensis*, which belongs to the Hynobiidae family and is more distantly related to Salamandridae than* A. mexicanum*, which belongs to the Ambystomatidae family [54, 55].*

References

1. Gregory, T.R., *Genome size and developmental complexity.* Genetica, 2002. **115**: p. 131-146.
2. Litvinchuk, S.N., J.M. Rosanov, and L.J. Borkin, *Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the Salamandridae (Urodela, Amphibia)*. Genome, 2007. **50**: p. 333-342.
3. Sessions, S.K., *Evolutionary cytogenetics in salamanders.* Chromosome Research, 2008. **16**(1): p. 183-201.
4. Sun, C., et al., *LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders.* Genome Biology and Evolution, 2012. **4**(2): p. 168-83.
5. Gregory, T.R. *Animal Genome Size Database*. 2024 14-11-2024]; Available from: http://www.genomesize.com.
6. Smith, J.J., et al., *A chromosome-scale assembly of the axolotl genome.* Genome Research, 2019. **29**: p. 317-324.
7. Lou, R.N., et al., *A beginner's guide to low-coverage whole genome sequencing for population genomics.* Molecular Ecology, 2021. **30**(23): p. 5966-5993.
8. Calboli, F.C., et al., *The need for jumpstarting amphibian genome projects.* Trends in Ecology & Evolution, 2011. **26**(8): p. 378-9.
9. Good, J.M., *Reduced representation methods for subgenomic enrichment and next-generation sequencing.* Molecular methods for evolutionary genetics, 2011: p. 85-103.
10. Zaharias, P., et al., *Data, time and money: evaluating the best compromise for inferring molecular phylogenies of non-model animal taxa*. Molecular Phylogenetics and Evolution, 2020. **142**: p. 106660.
11. Andermann, T., et al., *A guide to carrying out a phylogenomic target sequence capture project.* Frontiers in Genetics, 2019. **10**: p. 1407.
12. Da Fonseca, R.R., et al., *Next-generation biology: sequencing and data analysis approaches for non-model organisms.* Marine Genomics, 2016. **30**: p. 3-13.
13. Lemmon, E.M. and A.R. Lemmon, *High-throughput genomic data in systematics and phylogenetics.* Annual Review of Ecology, Evolution, and Systematics, 2013. **44**(1): p. 99-121.
14. Jones, M.R. and J.M. Good, *Targeted capture in evolutionary and ecological genomics.* Molecular Ecology, 2016. **25**(1): p. 185-202.
15. Hu, Q., et al., *Genome-wide RAD sequencing to identify a sex-specific marker in Chinese giant salamander Andrias davidianus.* BMC Genomics, 2019. **20**(1): p. 415.
16. Rancilhac, L., et al., *Phylogeny and species delimitation of near Eastern Neurergus newts (Salamandridae) based on genome-wide RADseq data analysis.* Molecular Phylogenetics and Evolution, 2019. **133**: p. 189-197.
17. Hubbs, N.W., et al., *Conservation genomics of urban populations of Streamside Salamander (Ambystoma barbouri).* PLoS ONE, 2022. **17**(6): p. e0260178.
18. Andrews, K.R., et al., *Harnessing the power of RADseq for ecological and evolutionary genomics.* Nature Reviews: Genetics, 2016. **17**(2): p. 81-92.
19. France, J., et al., *Linkage mapping vs Association: a comparison of two RADseq-based approaches to identify markers for homomorphic sex chromosomes in large genomes.* Authorea Preprints, 2024.
20. Babik, W., et al., *Limited evidence for genetic differentiation or adaptation in two amphibian species across replicated rural-urban gradients.* Evolutionary Applications, 2024. **17**(6): p. e13700.
21. Rodríguez, A., et al., *Inferring the shallow phylogeny of true salamanders (Salamandra) by multiple phylogenomic approaches.* Molecular Phylogenetics and Evolution, 2017. **115**: p. 16-26.
22. Burgon, J.D., et al., *Phylogenomic inference of species and subspecies diversity in the Palearctic salamander genus Salamandra.* Molecular Phylogenetics and Evolution, 2021. **157**: p. 107063.
23. Scott, P.A., et al., *Phylogenomics reveal species limits and inter-relationships in the narrow-range endemic lycian salamanders.* Molecular Phylogenetics and Evolution, 2024. **202**: p. 108205.
24. Arnold, B., et al., *RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling.* Molecular Ecology, 2013. **22**(11): p. 3179-90.
25. Rubin, B.E., R.H. Ree, and C.S. Moreau, *Inferring phylogenies from RAD sequence data.* PLoS ONE, 2012. **7**(4): p. e33394.

26. Davey, J.W., et al., *Special features of RAD Sequencing data: implications for genotyping.* Molecular Ecology, 2013. **22**(11): p. 3151-64.
27. Dodsworth, S., et al., *Hyb-seq for flowering plant systematics.* Trends in Plant Science, 2019. **24**(10): p. 887-891.
28. Lowry, D.B., et al., *Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation.* Molecular Ecology Resources, 2017. **17**(2): p. 142-152.
29. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing.* Nature Methods, 2010. **7**(2): p. 111-8.
30. Harvey, M.G., et al., *Sequence capture versus restriction site associated DNA sequencing for shallow systematics.* Systematic Biology, 2016. **65**: p. 910-924.
31. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.* Nature Biotechnology, 2009. **27**(2): p. 182-9.
32. Albert, T.J., et al., *Direct selection of human genomic loci by microarray hybridization.* Nature Methods, 2007. **4**(11): p. 903-5.
33. Grover, C.E., A. Salmon, and J.F. Wendel, *Targeted sequence capture as a powerful tool for evolutionary analysis.* American Journal of Botany, 2012. **99**(2): p. 312-9.
34. Heyduk, K., et al., *Targeted DNA Region Re-sequencing*, in *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing*. 2016. p. 43-68.
35. Fitzpatrick, B.M., et al., *New evidence contradicts the rapid spread of invasive genes into a threatened native species.* Biological Invasions, 2024.
36. Zhou, L. and J.A. Holliday, *Targeted enrichment of the black cottonwood (Populus trichocarpa) gene space using sequence capture.* BMC genomics, 2012. **13**: p. 1-12.
37. Featherstone, L.A. and A. McGaughran, *The effect of missing data on evolutionary analysis of sequence capture bycatch, with application to an agricultural pest.* Molecular Genetics and Genomics, 2024. **299**(1): p. 11.
38. Guo, Y., et al., *Exome sequencing generates high quality data in non-target regions.* BMC genomics, 2012. **13**: p. 1-10.
39. Bi, K., et al., *Transcriptome-based exon capture enables highly cost-effective comparative genomic data colection at moderate evolutionary scales.* BMC Genomics, 2012. **13**(403): p. 1471-2164.
40. Jimenez-Mena, B., et al., *Fishing for DNA? Designing baits for population genetics in target enrichment experiments: Guidelines, considerations and the new tool supeRbaits.* Molecular Ecology Resources, 2022. **22**(5): p. 2105-2119.
41. Khan, R., et al., *Development of a Target Enrichment Probe Set for Conifer (REMcon)*. Biology (Basel), 2024. **13**(6).
42. Quek, Z.B.R. and S.H. Ng, *Hybrid-capture target enrichment in human pathogens: identification, evolution, biosurveillance, and genomic epidemiology.* Pathogens, 2024. **13**(4).
43. Yu, P.L., et al., *Next-generation fungal identification using target enrichment and Nanopore sequencing.* BMC Genomics, 2023. **24**(1): p. 581.
44. Faircloth, B.C., et al., *Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera.* Molecular Ecology Resources, 2015. **15**(3): p. 489-501.
45. Teasdale, L.C., et al., *Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture.* Molecular Ecology Resources, 2016. **16**(5): p. 1107-23.
46. Schott, R.K., et al., *Targeted capture of complete coding regions across divergent species.* Genome Biology and Evolution, 2017. **9**(2): p. 398-414.
47. Singhal, S., et al., *Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles.* Molecular Ecology Resources, 2017. **17**(6): p. e12-e24.
48. Alfaro, M.E., et al., *Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary.* Nature Ecology & Evolution, 2018. **2**(4): p. 688-696.
49. Hutter, C.R., et al., *FrogCap: a modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales.* Molecular Ecology Resources, 2022. **22**(3): p. 1100-1119.
50. McCartney-Melstad, E., G.G. Mount, and H.B. Shaffer, *Exon capture optimization in amphibians with large genomes.* Molecular ecology resources, 2016. **16**: p. 1084-1094.

51. Wielstra, B., et al., *Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle*. Molecular Phylogenetics and Evolution, 2019. **133**: p. 120-127.
52. Bragg, J.G., et al., *Exon capture phylogenomics: efficacy across scales of divergence*. Molecular Ecology Resources, 2016. **16**(5): p. 1059-68.
53. Portik, D.M., L.L. Smith, and K. Bi, *An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura)*. Molecular Ecology Resources, 2016. **16**(5): p. 1069-83.
54. Marjanović, D. and M. Laurin, *An updated paleontological timetree of lissamphibians, with comments on the anatomy of Jurassic crown-group salamanders (Urodela)*. Historical Biology, 2013. **26**(4): p. 535-550.
55. Frost, D.R., *Amphibian species of the world: a taxonomic and geographic reference*. Vol. 1. 1985: Lawrence, KS: Association of Systematic Collections.
56. De Visser, M.C., et al., *Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying Triturus' balanced lethal system* bioRxiv, 2024: p. 2024.10.25.620277.
57. Kazilas, C., et al., *Spatial genetic structure in European marbled newts revealed with target enrichment by sequence capture*. Molecular Phylogenetics and Evolution, 2024. **194**: p. 108043.
58. Mars, J., et al., *Phylogenomics yields new systematic and taxonomical insights for Lissotriton newts, a lineage with a strong legacy of introgressive hybridization*. SSRN Preprints, 2024.
59. Kalaentzis, K., et al., *Phylogenomics resolves the ancient and rapid radiation of banded newts (genus Ommatotriton)*. SSRN Preprints, 2024.
60. Robbemont, J., et al., *An extended mtDNA phylogeography for the alpine newt illuminates the provenance of introduced populations*. Amphibia-Reptilia, 2023. **44**(3): p. 347-361.
61. Pasmans, F., et al., *Biogeography of Neurergus strauchii barani Öz, 1994 and N. s. strauchii (Steindachner, 1887) (Amphibia: Salamandridae) assessed using morphological and molecular data*. Amphibia-Reptilia, 2006. **27**(2): p. 281-288.
62. Valbuena-Urena, E., F. Amat, and S. Carranza, *Integrative phylogeography of Calotriton newts (Amphibia, Salamandridae), with special remarks on the conservation of the endangered Montseny brook newt (Calotriton arnoldi)*. PLoS ONE, 2013. **8**(6): p. e62542.
63. Sequeira, F., et al., *Genetic traces of hybrid zone movement across a fragmented habitat*. Journal of Evolutionary Biology, 2022. **35**(3): p. 400-412.
64. France, J., et al., *Genomic evidence suggests the balanced lethal system in Triturus newts originated in an instantaneous speciation event*. bioRxiv, 2024: p. 2024.10.29.620207.
65. Sambrook, J. and D.W. Russell, *Molecular cloning: a laboratory manual*. 2001, Coldspring-Harbour Laboratory Press: London, United Kingdom.
66. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data*. Bioinformatics, 2014. **30**(15): p. 2114-20.
67. Andrews, S., *FastQC: A Quality Control Tool for High Throughput Sequence Data*. 2010.
68. Andermann, T., et al., *SECAPR-a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments*. PeerJ, 2018. **6**: p. e5175.
69. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv, 2013(1303.3997).
70. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
71. Danecek, P., et al., *Twelve years of SAMtools and BCFtools*. Gigascience, 2021. **10**(2).
72. Auwera, G.A. and B.D. O'Connor, *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. 2020: O'Reilly Media.
73. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data*. Bioinformatics, 2011. **27**(21): p. 2987-93.
74. Rancilhac, L., et al., *Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders*. Molecular Phylogenetics and Evolution, 2021. **155**.
75. Poplin, R., et al., *Scaling accurate genetic variant discovery to tens of thousands of samples*. bioRxiv, 2017. **201178**.
76. Ortiz, E.M., *vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis*. 2019.

77. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*. Bioinformatics, 2014. **30**(9): p. 1312-3.
78. Lewis, P.O., *A likelihood approach to estimating phylogeny from discrete morphological character data*. Systematic biology, 2001. **50**(6): p. 913-925.
79. Wiens, B.J. and J.P. Colella, *triangulaR: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones*. bioRxiv, 2024: p. 2024.03.28.587167.
80. Wickham, H., *ggplot2*. Wiley interdisciplinary reviews: computational statistics, 2011. **3**(2): p. 180-185.
81. Knaus, B.J. and N.J. Grunwald, *vcfr: a package to manipulate and visualize variant call format data in R*. Molecular Ecology Resources, 2017. **17**(1): p. 44-53.
82. Zheng, X., et al., *A high-performance computing toolset for relatedness and principal component analysis of SNP data*. Bioinformatics, 2012. **28**(24): p. 3326-8.
83. Wielstra, B., et al., *A signature of dynamic biogeography: enclaves indicate past species replacement*. Proceedings of the Royal Society B: Biological Sciences, 2017. **284**(1868).
84. Team_R_Studio, *RStudio: integrated development environment for R*. 2021.
85. Team_R_Core, *RA language and environment for statistical computing, R Foundation for Statistical*. Computing, 2020.
86. Schilbert, H.M., A. Rempel, and B. Pucker, *Comparison of Read Mapping and Variant Calling Tools for the Analysis of Plant NGS Data*. Plants (Basel), 2020. **9**(4).
87. Steinfartz, S., et al., *A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary atterns of the Salamandridae with emphasis on Triturus newts*. Journal of experimental zoology. Part B, Molecular and developmental evolution, 2007. **308B**: p. 139-162.
88. Mirarab, S., et al., *A region of suppressed recombination misleads neoavian phylogenomics*. Proceedings of the National Academy of Sciences, 2024. **121**(15): p. e2319506121.
89. Zhang, D., et al., *Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow*. Systematic Biology, 2021. **70**(5): p. 961-975.
90. Veith, M., et al., *The changing views on the evolutionary relationships of extant Salamandridae (Amphibia: Urodela)*. PLoS ONE, 2018. **13**: p. 1-16.
91. Zhang, P., et al., *Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes*. Molecular Phylogenetics and Evolution, 2008. **49**(2): p. 586-97.
92. Kieren, S., et al., *A biogeographic and ecological perspective to the evolution of reproductive behaviour in the family Salamandridae*. Molecular Phylogenetics and Evolution, 2018. **121**: p. 98-109.
93. Lawson, G.R. and E.S. Kilpatrick, *Phylogeographic patterns among the subspecies of Notophthalmus viridescens (Eastern newt) in South Carolina*. Southeastern Naturalist, 2014. **13**(3): p. 444-455.
94. Gabor, C.R. and C.C. Nice, *Genetic variation among populations of eastern newts, Notophthalmus viridescens: a preliminary analysis based on allozymes*. Herpetologica, 2004. **60**(3): p. 373-386.
95. Whitmore, S.S., et al., *Conservation genetics of the central newt (Notophthalmus viridescens) in Iowa: the importance of a biogeographic framework*. Conservation Genetics, 2013. **14**(4): p. 771-781.
96. Kuchta, S.R. and A.-M. Tan, *Lineage diversification on an evolving landscape: phylogeography of the California newt, Taricha torosa (Caudata: Salamandridae)*. Biological Journal of the Linnean Society, 2006. **89**(2): p. 213-239.
97. Kuchta, S.R. and A.M. Tan, *Isolation by distance and post-glacial range expansion in the rough-skinned newt, Taricha granulosa*. Molecular Ecology, 2005. **14**(1): p. 225-44.
98. Tan, A.-M. and D.B. Wake, *MtDNA phylogeography of the California newt, Taricha torosa (Caudata, Salamandridae)*. Molecular Phylogenetics and Evolution, 1995. **4**(4): p. 383-394.
99. Tan, A.-M., *Systematics, phylogeny and biogeography of the northwest American newts of the genus Taricha (Caudata: Salamandridae)*. 1993: University of California, Berkeley.
100. Dufresnes, C. and A. Hernandez, *Towards completing the crocodile newts' puzzle with all-inclusive phylogeographic resources*. Zoological Journal of the Linnean Society, 2022. **197**(3): p. 620-640.
101. Weisrock, D.W., et al., *A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata)*. Molecular Phylogenetics and Evolution, 2006. **41**(2): p. 368-83.
102. Mattoccia, M., et al., *Phylogeography of an Italian endemic salamander (genus Salamandrina): glacial refugia, postglacial expansions, and secondary contact*. Biological Journal of the Linnean Society, 2011. **104**(4): p. 903-992.

103. Mattoccia, M., A. Romano, and V. Sbordoni, *Mitochondrial DNA sequence analysis of the spectacled salamander, Salamandrina terdigitata (Urodela: Salamandridae), supports the existence of two distinct species.* Zootaxa, 2005. **995**(1): p. 1–19-1–19.
104. Canestrelli, D., R. Bisconti, and G. Nascetti, *Extensive unidirectional introgression between two salamander lineages of ancient divergence and its evolutionary implications.* Scientific Reports, 2014. **4**: p. 6516.
105. Hauswaldt, J.S., et al., *From species divergence to population structure: a multimarker approach on the most basal lineage of Salamandridae, the spectacled salamanders (genus Salamandrina) from Italy.* Molecular Phylogenetics and Evolution, 2014. **70**: p. 1-12.
106. Veith, M., et al., *A molecular phylogeny of 'true' salamanders (family Salamandridae) and the evolution of terrestriality of reproductive modes.* Journal of Zoological Systematics and Evolutionary Research, 2009. **36**(1-2): p. 7-16.
107. IUCN_SSC_Amphibian_Specialist_Group. *Chioglossa lusitanica. The IUCN Red List of Threatened Species 2022*. e.T4657A89698017 2022; Available from: https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T4657A89698017.en.
108. Liu, L., et al., *Genetic consequences of long-term small effective population size in the critically endangered pygmy hog.* Evolutionary Applications, 2021. **14**(3): p. 710-720.
109. Vane-Wright, R.I., C.J. Humphries, and P.H. Williams, *What to protect?—Systematics and the agony of choice.* Biological conservation, 1991. **55**(3): p. 235-254.
110. Vitt, P., et al., *Global conservation prioritization for the Orchidaceae.* Scientific reports, 2023. **13**(1): p. 6718.
111. Tarkhnishvili, D.N., R.S. Thorpe, and J.W. Arntzen, *Pre-pleistocene refugia and differentiation between populations of the caucasian salamander (Mertensiella caucasica).* Molecular Phylogenetics and Evolution, 2000. **14**(3): p. 414-22.
112. Weisrock, D.W., et al., *Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in Mertensiella luschani associated with the rise of Anatolia.* Molecular Phylogenetics and Evolution, 2001. **18**(3): p. 434-448.
113. Fijarczyk, A., et al., *Balancing selection and introgression of newt immune-response genes.* Proceedings of the Royal Society B: Biological Sciences, 2018. **285**.
114. Li, G., et al., *Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae).* Genome Research, 2016. **26**: p. 1-11.
115. Mallet, J., *Hybridization as an invasion of the genome.* Trends in Ecology and Evolution, 2005. **20**: p. 229-237.
116. Quilodran, C.S., J.I. Montoya-Burgos, and M. Currat, *Harmonizing hybridization dissonance in conservation.* Communications Biology, 2020. **3**(1): p. 391.
117. Wielstra, B., et al., *Efficient screening for 'genetic pollution' in an anthropogenic crested newt hybrid zone.* Conservation Genetics Resources, 2016. **8**: p. 553-560.
118. Simberloff, D., *Invasive species: what everyone needs to know*. 2013: Oxford University Press.
119. Dufresnes, C., et al., *Introgressive hybridization of threatened European tree frogs (Hyla arborea) by introduced H. intermedia in Western Switzerland*. Conservation Genetics, 2015. **16**(6): p. 1507-1513.
120. Frankham, R., J.D. Ballou, and D.A. Briscoe, *A primer of conservation genetics*. 1st Editio ed. 2004, New York: Cambridge University Press.
121. Ørsted, M., et al., *Genomic variation predicts adaptive evolutionary responses better than population bottleneck history.* PLoS Genetics, 2019. **15**: p. 1-18.
122. Weisrock, D.W., et al., *Surmounting the large-genome "Problem" for genomic data generation in salamanders*, in *Population Genomics: Wildlife*, P.A. Hohenlohe and O.P. Rajora, Editors. 2018, Springer International Publishing: Cham. p. 115-142.
123. Stuart, S.N., et al., *Status and trends of amphibian declines and extinctions worldwide.* Science, 2004. **306**: p. 1783-1787.
124. Hayes, T.B., et al., *The cause of global amphibian declines: a developmental endocrinologist's perspective.* Journal of Experimental Biology, 2010. **213**: p. 921-933.
125. Sparreboom, M., *Salamanders of the Old World: the salamanders of Europe, Asia and northern Africa.* 2014: p. 431p.
126. Lucas, P.M., et al., *Using comparative extinction risk analysis to prioritize the IUCN Red List reassessments of amphibians.* Conservation Biology, 2024: p. e14316.

127. Martel, A., et al., *Batrachochytrium salamandrivorans sp. nov. causes lethal chytridiomycosis in amphibians*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**: p. 15325-15329.

Triturus *embryo breeding, Belgrade University, Serbia*

Chapter 5 - PAV-spotter: using signal cross-correlations to identify Presence/Absence Variation in target capture data

Manon de Visser[1,2], Chris van der Ploeg[3,4], Milena Cvijanović[5], Tijana Vučić[1,2,6], Anagnostis Theodoropoulos[1,2], Ben Wielstra[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands
2. Naturalis Biodiversity Center, Leiden, The Netherlands
3. Dynamics and Control Group, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
4. Integrated Vehicle Safety Group, Netherlands Organisation for Applied Scientific Research, Helmond, The Netherlands
5. Institute for Biological Research "Siniša Stanković", National Institute of the Republic of Serbia, Department of Evolutionary Biology, University of Belgrade, Belgrade, Serbia
6. Institute of Zoology, Faculty of Biology, University of Belgrade, Belgrade, Serbia

This manuscript has been submitted to the journal ***Molecular Ecology Resources*** and is currently under review. A preprinted version of this manuscript has been published at ***bioRxiv*** (2024.10.25.620064; https://doi.org/10.1101/2024.10.25.620064).

Abstract

High throughput sequencing technologies have become essential in the fields of evolutionary biology and genomics. When dealing with non-model organisms or genomic gigantism, sequencing whole genomes is still relatively costly and therefore reduced-genome representations are frequently obtained, for instance by 'target capture' approaches. While computational tools exist that can handle target capture data and identify small-scale variants such as single nucleotide polymorphisms and micro-indels, options to identify large scale structural variants are limited. To meet this need, we introduce PAV-spotter: a tool that can identify presence/absence variation (PAV) in target capture data. PAV-spotter conducts a signal cross-correlation calculation, in which the distribution of read counts per target between samples of different *a priori* defined classes – e.g. male versus female, or diseased versus healthy – are compared. We apply and test our methodology by studying *Triturus* newts: salamanders with gigantic genomes that currently lack an annotated reference genome. *Triturus* newts suffer from a hereditary disease that kills half their offspring during embryogenesis. We compare the target capture data of two different types of diseased embryos, characterized by unique deletions, with those of healthy embryos. Our findings show that PAV-spotter helps to expose such structural variants, even in the face of medium to low sequencing coverage levels, low sample sizes, and background noise due to mis-mapped reads. PAV-spotter can be used to study the structural variation underlying supergene systems in the absence of whole genome assemblies. The code, including further explanation, is available through the PAV-spotter GitHub repository: https://github.com/Wielstra-Lab/PAVspotter.

Keywords

PA variation, Hyb-Seq , target enrichment, copy number variation, pattern recognition, *Triturus*, supergenes, hemizygosity.

Introduction

Next-generation DNA sequencing methods have revolutionized the biological sciences, with an ever-growing amount of sequence data being generated worldwide [1, 2]. High throughput sequencing techniques have become more affordable and increasingly used, however sequencing entire genomes can still be challenging, for instance when dealing with non-model organisms, genomic gigantism, or a combination of the two [3-5]. In such cases, well-annotated reference genomes for aligning (re-)sequenced reads are generally unavailable, making whole genome sequencing relatively costly in terms of money and (computational) time [e.g. see; 6, 7, 8]. Many biologists therefore still opt for more cheap and efficient 'reduced-representation' high throughput sequencing techniques, which allow for a subset of loci to be sequenced more deeply [9-11].

A technique that has become particularly popular for studying non-model species is target capture, also referred to as hybridization sequencing, exon capture, sequence capture, or exome capture [11-14]. This method facilitates collecting sequence information on hundreds or thousands of pre-selected loci. Due to the consequent rise of large multi-locus DNA sequence datasets, the need for innovative, easy-to-implement bioinformatic applications has surged as well [15]. User-friendly pipelines help to pre-process sequence reads by wrapping and connecting existing software tools, with well-known examples for target capture including HybPiper [16], Assexon [17] and Sequence Capture Processor [i.e. SECAPR, see; 18]. These pipelines and their software dependencies support fast upstream data cleaning and guide the user to phylogenetic applications downstream.

Typically, target capture analysis pipelines are utilized to identify small-scale genetic variants such as single nucleotide polymorphisms (SNPs) and relatively small insertions/deletions (microindels, ranging from 1~50bp) to perform, for instance, phylogenetic tree-building [10]. However, to identify and analyze larger scale information from target capture data, such as genomic structural variation, few tools are available – especially when focusing on non-model organisms. A particular type of larger-scale variation that is hard to identify in multi-locus DNA sequence datasets is presence/absence variation (hereafter referred to as 'PAV') of relatively big insertions/deletions (macroindels, > 50bp), i.e. above the size of microindels [19, 20].

Structural variants such as PAVs are regularly overlooked because they are harder to identify than SNPs [21, 22]. However they are a major source of genetic divergence and diversity [23-26]. PAV in particular poses an extreme example of copy number variation, where fragments in the size of entire exons or (stretches of DNA containing multiple) genes are missing from one genome compared to another [27, 28]. When comparing such genomes, target capture data would in theory display PAV by

showing 'normal data' in the case of target presence, versus a 'data gap' in the case of target absence. This would be an indication of structural variation.

Whether there are consistent differences in presence or absence of sequence data can be determined by analyzing the way that reads pile up against a certain reference set of sequences. Some tools can detect copy number variation and PAV patterns by comparing the depth of mapped reads of different samples, such as ExomeCNV (Sathirapongsasuti et al, 2011) and SUPER-CAP (Yuan et al, 2019). However, these tools come with strict requirements, such as good quality reference genomes being available, known functional annotation of variants, and/or coverage being consistently high across all samples and targets, with mapped reads ideally following a normal distribution. Yet, most multi-locus datasets, including those resulting from target capture experiments, generally do not meet such requirements [10].

We introduce PAV-spotter: a flexible signal cross-correlation method that is able to 'spot' potential PAV in target capture datasets. Our approach borrows the notion of cross-correlation to detect the dissimilarity between datasets obtained through target capture experiments. Cross-correlation methods are generally used in the domain of control engineering. Classically, they are applied on time-series data [29], for example in machinery fault detection studies [30]. However, cross-correlation approaches have been proven useful in the field of pattern recognition as well [31].

Being able to identify structural variation by using pattern recognition would be especially informative when studying supergenes systems, in which individuals can have zero, one or two copies of particular loci [32, 33]. Supergenes consist of genes that are inherited together as a single locus due to the suppression of recombination [33-35]. As a result, the non-recombining stretches of 'supergene DNA' evolve independently of one another, facilitating the rapid evolution of complex adaptations [36, 37]. These sets of genes are often polymorphic and subjected to balancing selection, as a species generally possesses at least two supergene variants [38]. Sex chromosomes, for instance, are classically considered supergenes [39]. In diploid organisms the heterogametic sex inherits the sex-determining 'supergene', as well as the alternate sex chromosome, in a hemizygous manner – meaning that they only receive one copy of each [40]. In the XX-XY sex determination system of mammals, for instance, males generally possess a single copy of the supergene that is the Y chromosome (as well as a single copy of the X-chromosome), whereas in the ZW-WW system it is the females that possess the Z supergene once (next to a single W chromosome). Hence, genes that are hemizygous and thus lie solely on the sex-determining supergene would show PAV in target capture datasets when data of different sexes is compared.

However, supergene systems are not limited to the biological concept of sex. Other, famous examples of supergenes underlying complex traits are; the Müllerian mimicry complex in *Numata* longwing butterflies [39], the striking sexual dimorphism and

breeding behaviors of ruffs [41, 42] and white-throated sparrows [43], the social polymorphism observed in several species of ant [44], and heterostyly in primrose flowers [45]. Furthermore, hemizygous inheritance of (super)genes also occurs in, for instance, genetic incompatibilities such as with "hybrid necrosis" in plants, which can be linked to PAV in certain genes in for example Asian rice [46, 47]; hereditary diseases such as α-thalassaemia, which is caused by large deletions in the alpha globin gene cluster on chromosome 16 in humans [48]; and in balanced lethal systems, in which two distinct chromosome forms exist that are covered by unique lethal mutations [49].

We demonstrate the application of PAV-spotter using the balanced lethal system in *Triturus* newts as a case study. *Triturus* individuals either are heteromorphic and possess two different versions of their largest autosomal chromosome, characterized by unique deletions, or they are homomorphic and possess two identical versions of this chromosome [50]. The two types of homomorphic individuals express a unique disease state and both die during embryogenesis, whereas the heteromorphic individuals are viable [51-54]. This 'double hemizygous' system lends itself particularly well for using target capture data to detect PAV, as it allows for a reciprocal test: targets deleted from one chromosome version should be present on the alternate version and the other way around. Based on our findings, we describe the usefulness, as well as the limitations, of our approach.

Methods

Sample information & collection

The first chromosome of *Triturus* comes in two forms: 1A and 1B. Homomorphic individuals (1A1A or 1B1B) invariably die during embryogenesis, while heteromorphic individuals (1A1B/1B1A) are viable. We collected *T. macedonicus* x *T. ivanbureschi* F_1 hybrid embryos from an ongoing breeding experiment at the Institute for biological research, „Siniša Stanković", University of Belgrade [with experimental settings, breeding conditions, and other details on the process of raising embryos as described in; 55, 56]. Embryo development was followed through observation with a stereomicroscope. Diseased embryos were collected when the process leading up to developmental arrest occurred, which is visible as a 'growth slowdown', during the late tail-bud phase. Diseased embryos were then classified into either the "fat-tailed" (FT) phenotype or the "slim-tailed" (ST) phenotype based on morphological characteristics of the embryo [53, 57]. Healthy/control (HC) embryos that survived this critical phase were subsequently collected. We collected 30 individuals in total (Supplementary Table 1); ten of each class, i.e. ten ST, ten FT and ten HC embryos. Samples were stored in ethanol at -20 °C until further handling.

Laboratory procedures & pre-processing of sequence data

We followed the standard "NewtCap" workflow of salt-based extraction of DNA from embryonic tissue, followed by quantification, library preparation, target capture and Illumina sequencing [as described in; 58]. After obtaining the raw, paired-end sequence reads from Baseclear B.V. (Leiden, the Netherlands), we followed a standard pipeline for checking the quality of, and for cleaning-up and mapping, our sequence data in a Linux environment up to and until the deduplication of the BAM files step [as described in; 58]. These deduplicated BAM files served as input for further data extraction and analyses. Throughout the cleaning and mapping process, we used SAMtools' [59] *stats*, *flagstat* and *coverage* options to calculate basic statistics from the FASTQ and BAM files. The reference FASTA file used for read mapping can be found in Supplementary Material as 'Targets.fasta'. These 7,139 sequences, initially used for probe tiling, were based on *T. dobrogicus* transcripts, and had a maximum length of 450bp [58, 60].

Preparing read depth data for PAV-spotter

From the BAM files we extracted information on sequence read depth for all sites per target by using the SAMtools depth option [59]. We optimized this extracted information by following several file-manipulation steps in a custom 'prepping' shell script (Script 1) to make the input files and folder structure match the requirements of our PAV-spotter tool. The steps in Script 1 include automatically merging and sorting of intermediate files where appropriate, changing the tab-delimited format to a CSV format, splitting the overall CSV file into multiple files (one file per separate target/gene), and creating a text file with sample names for later use (details are explained in the script). The exact format of the input folder structure and input files is described in Box 1.

PAV-spotter assumes background knowledge on phenotype classes that presumably differ in the presence of certain genes (in other words, cross-comparisons are not random: for instance males are compared to females, diseased individuals are compared to control samples, etc.). Here, we work with the a priori classification of three phenotypes: two types of diseased embryos (FT vs. ST) and healthy embryos to serve as a control (HC). Which sample belongs to which class needs to be designated by your input filenames (the BAM files), which will end up in an automatically created text file "individuals.txt" after running Script 1. This is crucial, as PAV-spotter performs the comparisons based on the phenotype information embedded in the names of this text file. In case phenotypic classes as specified by the user cannot be deduced from input file names, the user needs to either alter the input file names manually, or alter the identifiers in Script 1 manually – or both – before running Script 1 (ideally, the user includes such filename identifiers at the raw FASTQ file stage for consistency throughout the pipeline).

Applying PAV-spotter

We applied Script 1 on our total set of 30 samples (Supplementary Table 2; n=30, ten 'ST', ten 'FT' and ten 'HC' individuals, with these identifier abbreviations occurring in the sample names). Additionally, we applied the script on random subsets of samples (Supplementary Table 3 and Supplementary Table 4); two separate analyses with a sample size of five per class (i.e. two total subsets of n=15, indicated by sample set '5_1' and run '5_2'), and five more separate analyses with a sample size of two per class (i.e. five total subsets of n=6, indicated by sample set '2_1', '2_2', '2_3', '2_4' and '2_5'). This allowed us to assess the performance of PAV-spotter when lower sample sizes are used. We randomized the grouping of samples into subsets by using the 'shuf' command from the standard GNU Core Utilities (http://gnu.org/s/coreutils/).

We ran a custom MATLAB script (Script 2, hereafter 'PAV-spotter') remotely through SLURM workload manager (example batch script attached). Users can compare two, or three phenotypic classes (using the argument "categories", see the SLURM script), but if only two are provided, only two are compared. Also, we indicated which class we considered the control group (by implementing "ctrl_category = 'HC'") and we provided a common identifier for all input files/targets (with the argument "common_identifier = 'DN'"). Also, information on the working directory and desired output filename was provided in the MATLAB command depending on the analysis, and the same is the case for the customized #SBATCH lines for running the SLURM job.

BOX 1:
This is a description of the expected input file format for the PAV-spotter script:

'Species' Directories:
- Each species, or otherwise distinguishable set of data, has its own directory
- PAV-spotter is built in such a way that it will loop over multiple such directories

"individuals.txt" file:
- An automatically generated file, uses the initial sample names for input
- Needs to be located in, and corresponding to the contents of, a particular species directory
- This file contains the individual information, with each sample name on a new line and with information on the classes to be compared included in the name (e.g. 'ST', 'FT', and 'HC')

Gene/Target Data Files:
- Also needs to be located in, and corresponding to the contents of, a particular species directory
- Each file represents a single gene or target
- Each file has columns and is in CSV format (this should be the output from batch script 1):
    - Column 1: Gene/target name
    - Column 2: Gene/target position (a number)
    - Column 3: Read depth data (a number)
    - Column 4: Sample/class name (should match identifiers in "individuals.txt" file)

The script was run four times on our total set of n=30: once with default settings (no filtering, “reads_threshold” == 0 and “contig_width” == 0), one time with a mild coverage filtering (“reads_threshold” == 5 and “contig_width” == 0), one time with a mild filtering for minimum length of contigs (“reads_threshold” == 0, “contig_width” == 50) and one time with both of the filtering thresholds (“reads_threshold” == 5 and “contig_width” == 50). By setting a soft coverage filter of a minimum of five reads, we filter out reads of any poorly covered target of an individual that does not meet this criterium (thereby assuming a coverage of zero across the target in question). Furthermore, by specifying a minimum contig width, the script will filter out read information of covered regions in between two positions with zero coverage, in case those regions are narrower than, in this case, 50bp (thereby assuming a coverage of zero in that specific target region). For the tests on sample subsets of five individuals per class and two individuals per class, we only used the default filtering settings (no filtering, “reads_threshold” == 0 and “contig_width” == 0). In all analyses, we enabled the script to plot accessory figures (“plot_figures” == TRUE, clarification below). MATLAB v.9.13.0 (https://www.mathworks.com) was used for running PAV-spotter.

The rationale behind PAV-spotter

After setting up the input files and initiating PAV-spotter successfully, filtering settings are applied as specified. Subsequently, PAV-spotter merges and normalizes the sequence read depth information for all the samples per target per specified sampling class before it starts the actual cross-correlation analyses (but users can turn this setting off in case calculations of all possible cross-comparisons on an individual level are desired). By exploiting the availability of multiple samples per phenotypic class in this way, we ensure that the comparisons will be made on a class level rather than on an individual level. In the latter case, more false outcomes would be expected as a result of the stochastic nature of target capture experiments, something that can be avoided by pooling results. PAV-spotter can loop through a set of input directories if separate datasets need to be analyzed with similar settings consecutively

The cross-correlation analysis in PAV-spotter works as follows (Figure 1). Two target capture results of the same targeted region, but of a different phenotypic class, are defined by D_1 and D_2. These represent two vectors of identical length T, where the vector position represents the target position and the vector values represent the normalized sequence read depth. PAV-spotter then calculates the similarity of the cross-correlation formulated by

$$C_{1,2}[n] = \sum_{m=-T}^{T} D_1[m] D_2[m+n]$$

for all n between $-T$ and T, with n and m accessing the indices of the vectors $C_{1,2}, D_1, D_2$. This constructs a cross-correlation vector of which the maximum value provides the maximum similarity of D_1 and D_2. For example, in the case that $D_1 = D_2$ (i.e., in the case of autocorrelation), the similarity score C will be 100%.

PAV-spotter outputs a CSV file per separate analysis with all the cross-correlation data in the form of percentage similarity, and it outputs a folder with figures. Also, a CSV file with cross-comparisons between data from samples from the control group only (the HCs) is generated. This extra file allows for a calculation of the overall resemblance of the control samples, which should have data present with well-captured targets. This as opposed to the FT and ST samples, which are expected to show 'data gaps', or absence, for some targets.

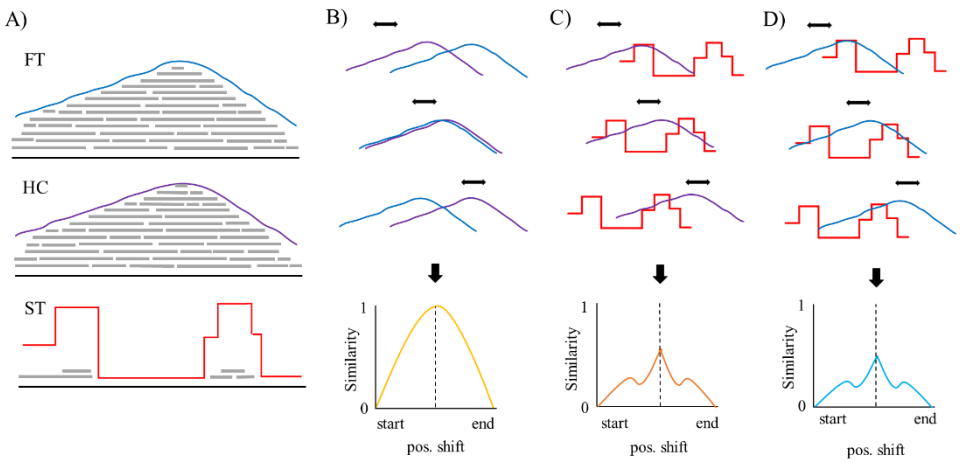


Figure 1: *A simplified visualization of the cross-correlation methodology in PAV-spotter, showing a hypothetical gene/target that is missing in only ST embryos as an example.* ***(A)*** *A number of sequence reads are mapped (in grey) against a gene/target. This information is taken from depth files and merged per phenotypic class (in case of multiple samples per class), then the absolute distributions of the read depths are normalized, here represented by the colored lines; dark blue = FT, purple = HC, red = ST.* ***(B)*** *The distribution data of the FT and the HC classes are compared. A measure of similarity is determined as a function of the displacement of one read depth distribution relative to the other, as if they were to 'slide over' each other (indicated by black arrows). The similarity appears close to 1 (=100%) and the graph produced by PAV-spotter also follows a smooth line.* ***(C)*** *The distribution data of the HC and ST classes are compared: the similarity is not close to 100% and the similarity graph produced by PAV-spotter does not follow a smooth line.* ***(D)*** *The distribution data of the HC and ST classes are compared: the similarity is not close to 100% and the similarity graph produced by PAV-spotter does not follow a smooth line.*

As we are investigating a double hemizygous system, we always have three cross-correlation values to work with. This means we are able to use not only the healthy embryos (HC), but also the other class of diseased embryo (ST or FT), as a control, because genes absent in one class of diseased embryo are expected to be present in both other embryo classes (e.g. to recognize absence in ST embryos, which should have a 1A1A genotype, we can check for presence in the HC embryos which should have the 1A1B genotype, but we can do an additional check for presence in the ST embryos that should have a 1B1B genotype – and the same applies the other way around). Hence, to deduce PAV in the chromosome that is inherited twice in ST embryos, we search for a pattern in which a significant portion of the target was present in both HC and FT embryos (which contain the alternate chromosome form), but absent in ST embryos. Conversely, to deduce PAV in the chromosome inherited twice by FT embryos, we searched for absence in FT, but presence in both HC and ST embryos.

Downstream PAV estimation

To automatically deduce PAV patterns, we applied another custom shell script, Script 3. This script takes the main output file of PAV-spotter, creates an overall matrix of the results, adds columns with information on the cross-correlation data that stand out, and makes lists of the targets that show potential PAV based on a threshold (which can be customized). When, for a certain target, the data of FT versus ST embryos were less than 80% similar, the data of FT versus HC embryos were less than 80% similar, and the data of ST versus HC embryos were more than 80% similar, this target was scored as '1A-linked'. For '1B-linked' genes it was the other way around: FT vs. ST embryos < 80% similar, ST vs. HC embryos < 80% similar, and FT vs ST embryos > 80% similar. This threshold of 20% dissimilarity equals a p-value of between 0.01 (≈ 15% dissimilarity) and 0.001 (≈ 28% dissimilarity). The PAV-spotter output file that shows all similarity scores among the HC embryos only, guided our choice for this threshold (e.g. see Figures 1 and 2).

Finally, as not much is known about the genetic background of our non-model study species *Triturus*, we performed visual inspections of the read content of all BAM files (n=30) that were used as input for PAV-spotter by checking them in Integrative Genomics Viewer (IGV) software (Robinson et al, 2011) on a Windows environment. This constitutes the last step in our overall workflow (Figure 2). We automated obtaining screenshots through IGV by running batch commands for all 7,139 targets (an example batch script is available). With a 'checking-by-eye' approach we categorized PAV hits as 'likely true' and 'likely false' in order to assess the quantity of false positive outcomes and we cross-checked the results of all the different runs to identify false negative outcomes (in other words: in case a 'likely true' target with PAV was retrieved in one analysis, but not in another, we counted it as a false negative in the latter analysis).

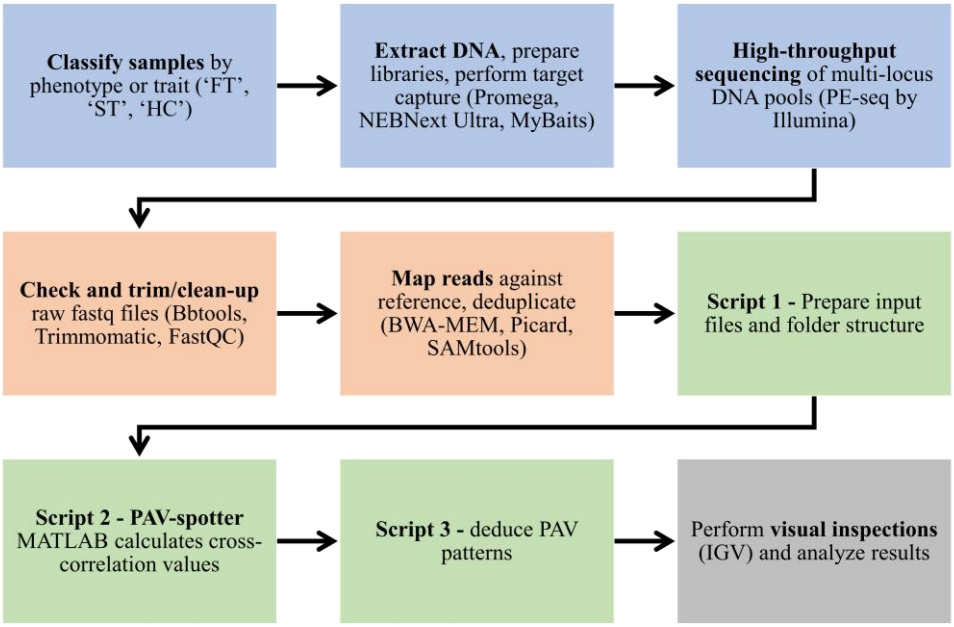


Figure 2: *A summary of the consecutive steps of our methods in which the order of the steps is indicated by black arrows. Blue boxes show the laboratory process, orange boxes represent bioinformatic pre-processing steps, green boxes stand for the application of the main PAV-spotter scripts, and the grey box covers the conclusive steps of inspecting and interpreting the results.*

We executed all bioinformatic steps, from pre-processing of reads and read depth information to applying PAV-spotter and extracting information from the output, through the High Performance Computing facility called 'ALICE' (Academic Leiden Interdisciplinary Cluster Environment, the Netherlands). The GitHub repository of PAV-spotter provides all the scripts and further explanation: https://github.com/Wielstra-Lab/PAVspotter.

Results

A mean of 6,483,062 read pairs were generated on average per sample, with a standard deviation of 1,450,648 read pairs (Supplementary Table 1). After trimming, this changed to a total of 6,187,680 read pairs with a SD of 1,369,254 read pairs. On average, 35.57% of the trimmed reads were successfully mapped against the reference targets after duplicate removal, as an average of 17.09% of all trimmed reads were flagged as duplicates (Supplementary Table 1).

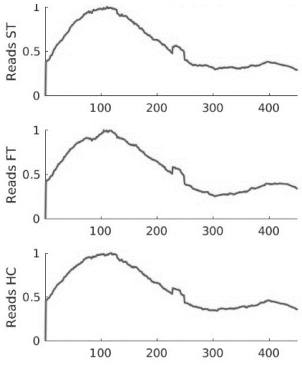
Overall, the targets had a mean read depth of 90.09 sequences and a mean coverage of 97.19 % of the sequence bases (Supplementary Table 2, presented per phenotypic class). For the overall set with ten samples per phenotypic class, the average depth of coverage was 84.7 in the FT group, 97.7 in the group of HC embryos, and 87.8 in the ST embryo group (Supplementary Table 2). Moreover, for the batched samples with five individuals per phenotypic class, this average depth of coverage varied between the lowest number of 76.6 in FT batch 5-2 and the highest number of 102.2 in HC batch 5-1 (Supplementary Table 3). Lastly, for the batched samples with two individuals per phenotypic class, the averages varied between the lowest number of 37.1 in FT batch 2-1, and the highest number of 134.1 in FT batch 2-5 (Supplementary Table 4).

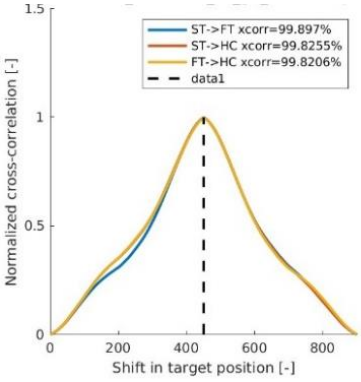
Through the different runs with a sample size of ten per class, we discovered large-scale PAV for in total 72 targets. Genes without PAV had sequence reads with similar distributions for all three classes (Figure 3A), whereas of those 72 aberrant genes that we discovered, 32 showed an absence in FT, but a presence in HC and ST embryos (and are thus "1A-linked", Table 5 and Figure 3B), while the remaining 40 were absent in ST, but not in HC and FT embryos (and are thus "1B-linked", Table 6 and Figure 3C). We confirmed our findings using the visual IGV inspection (for examples, see Supplementary Figures 3-5).

After running PAV-spotter without any filtering options on the full dataset, we correctly identified all 32 1A targets and generated one false positive in the 1A list, which we identified by the visual IGV inspection (see Supplementary Figure 6). By re-running PAV-spotter with the previously described filtering options, this false positive was removed from the list in some cases, however these additional analyses also generated more false positives and false negatives: the frequency of them depending on the combination of filtering settings used (Supplementary Table 5). For the 40 1B targets, we discovered 37 true positives and three false positives in the unfiltered run. Depending on the combination of filtering settings applied, two of these three false positives again disappeared from the list. However, these extra runs with filtering settings also highlighted three additional 1B targets that were overlooked (as false negatives) in the initial analysis (Supplementary Table 6).

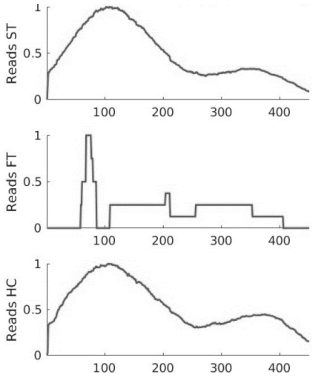
The false positive outcomes consistently had either the lowest - or in a single case, the highest - mean depth values, reflected by the 'MAXpeak' output of the overall results matrix generated by PAV-spotter. This value is the peak number of reads in one position of a certain target observed across all the individual samples included in the analyses. The false negative outcomes that came to light as true positive results after additional filtering was applied, all had similarity scores for the resembling classes (FT and HC) above 99% in the unfiltered analysis. But the lower similarity scores of these targets between the non-resembling classes (both between FT and ST and between ST and HC) were lower than 90%, but not lower than the 80% threshold (which is why they were initially overlooked).

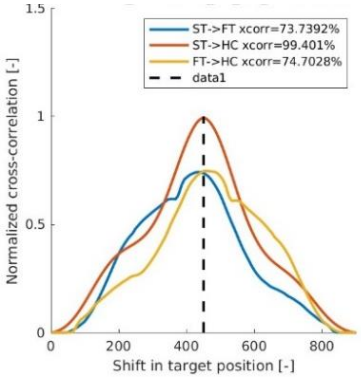
A) CDK



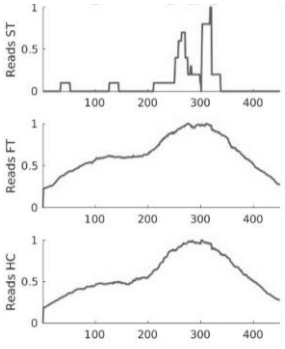


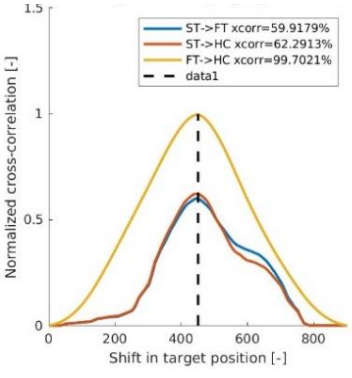
B) PLEKHM1





C) NAGLU





◄ ***Figure 3:*** *Examples of target-specific plots produced by PAV-spotter in the overall run with no filtering (n=10 per class). The stacked plots on the left side within the panels A-C show the merged and normalized distributions of sequence read depths per phenotypic class, and the colored plots on the right side within the panels A-C show associated measures of similarity of those distributions as a function of the displacement of one relative to the other (including a legenda explaining the colors and cross-correlation values).* ***A)*** *An example of a 'normal' gene/target, showing the cross-correlation analyses of control marker 'CDK' (see Discussion), with a similar shape of the sequence read depth distributions and high correlation values (above the 80% similarity threshold) between all three classes.* ***B)*** *An example of a 1A-linked gene/target, showing the cross-correlation analyses of 'PLEKHM1' (see Discussion), with a deviant read depth distribution for the FT class and a high correlation value (>80%) for ST vs. HC samples, but a lower cross-correlation value (<80%) for the ST vs. FT and FT vs. HC sample comparisons.* ***C)*** *An example of a 1B-linked gene/target, showing the cross-correlation analyses of 'NAGLU' (see Discussion), with a deviant read depth distribution for the ST class and a high correlation value (>80%) for FT vs. HC samples, but lower cross-correlation values (<80%) for the FT vs. ST and the ST vs. HC samples.*

The results for the n=5 per class runs (5_1 and 5_2) resemble our earlier findings. We re-discovered an average of 68.5 (95.1% success) out of the total of 72 PAV targets previously discovered, but with half the sample size. For run 5_1, this number was 70 out of 72 PAV targets (97.2% success) and for run 5_2 this number was 67 out of 72 (93.1% success). Overall, for 1A, the results of both the analyses with a sample size of five individuals per class were complementary, as all 32 previously identified 1A targets were re-discovered at least once (Supplementary Table 7). The same goes for the 40 previously identified 1B targets (Supplementary Table 8). Between the two runs, the overall mean depth of coverage was the lowest in 5_2, the analyses that also showed less successful out of the two.

The analysis of the 5_1 subset (n=15) resulted in the discovery of 31 true 1A targets plus one false negative and one two positives, and in 39 true 1B targets with one false negatives and one false positive. The analysis of the 5_2 subset (n=15) again resulted in 31 true 1A targets (with one difference) plus one false negative and two false positives. These false outcomes were not the same as with the 5_1 subset analysis. For 1B, the 5_2 analysis yielded 36 true 1B targets with four false negatives and seven false positives.

The results of the n=2 per class runs (2_1 through 2_5, each with n=6 in total) again highlighted the same PAV exhibiting targets. We re-discovered an average of 70 (97.2% success) out of the total of 72 PAV targets previously discovered, but with a fifth of the sample size (Supplementary Tables 9 and 10). For each of the 2_1, 2_2 and 2_4 runs, these numbers were indeed 70 out of 72 (97.2% success), run 2_3 retrieved 71 out of 72 PAV targets (98.6% success) and run 2_5 69 out of 72 (95.8%).

Overall, for both the 32 previously identified 1A targets and the 40 previously identified 1B targets, these five analyses with a sample size of only two individuals per class appeared complementary, as all true positive targets were re-discovered at least once. For 1A, one of the false negative outcomes came forward as a false negative in two out of the five analyses (and as a true positive in the three other analyses). The other three false negative outcomes in the 1A list were incidental. For 1B, each of the false negative outcomes occurred in only one of the five analyses (and formed a true positive result in the four alternative analyses). The number of false positive outcomes was slightly higher with these low sample size tests (Supplementary Tables 11 and 12), however this was especially noticeable for the 1A results of the first batch (2_1). The mean depth of coverage was also the lowest for the FT (1B1B) samples in this batch (Supplementary Table 4). We therefore tested for a correlation and show that the mean depth of the samples exhibiting absence (i.e. the mean depth of FT samples with determining 1A absence, and the mean depth of the ST with determining 1B absence) appeared to be negatively and significantly correlated to the number of false positives brought forward (Spearmann's rank correlation, n=10, p=0.018).

Discussion

We employ a signal cross-correlation approach to discern PAV patterns in target capture data of *Triturus* newt DNA. By comparing the read depth in sequence data of embryos of different phenotypic classes, and by manually checking the results of the read alignments, we are able to identify over seventy targets that appear to be either present in, or absent from, the genome, depending on the phenotype.

The three example targets displayed in Figure 3 have been independently tested using multiplex (mx) PCR techniques, including mxKASP [61]. Control marker CDK is present in all three embryo classes. This corresponds with our results from PAV-spotter, where we discover high cross-correlations values between the distribution of mapped reads of all three phenotypic classes for this target marker (>99% similarity, way above our 80% cutoff threshold). On the other hand, PLEKHM1 is observed in HC and ST embryos, but not in FT embryos, and NAGLU is observed in HC and FT embryos, but not in ST embryos [61]. Again, this matches our PAV-spotter findings – even with sample sizes as low as two individuals per phenotypic class.

Evidently, PAV-spotter relies on the correct, a priori classification of samples. Also, it merges the read depth data of samples per each phenotypic class in case multiple samples are provided (a setting that is recommended, but can be turned off if desired). Thus, future users should carefully consider what they want to compare. Additionally, we

underline that cross-contamination of DNA is, for example, a main concern with target capture of ancient DNA [62] and it could potentially distort the similarity values calculated by PAV-spotter – something to be wary of.

The fact that PAV-spotter is able to, on average, re-discover 97,2% of true positive PAV target outcomes in our trials is especially convenient for studies where scientists must rely on a limited amount of available DNA, as is often the case with herbarium specimens [63, 64]. However, the highest yield of false positive outcomes is observed in our 1A results of batch 2_1, but not the accessory 1B results, which firstly shows that there is a likely trade-off between sample size and sequence coverage. Preferably, the quality of DNA is as high as possible when working with small sample sizes. Although also preferred in case of a larger sample size per phenotypic class, the chances are then higher that any poor coverage sample(s) will be compensated for by sample(s) with better coverage. In general, spreading sequencing efforts across at least a couple samples, with slightly lower – but still informative – depths per sample, is considered a safer option than working with extremely low sample sizes [65, 66].

Regardless of sample size, identifying and characterizing structural variation from target capture data is widely recognized to be difficult [67] and it is especially challenging with non-model organisms. This is because capture-rates may vary considerably depending on bait design, sample quality, species relatedness, batch effects, and stochastic factors [10, 68]. In most cases where targets showed a significant absence of mapped reads, we observe a small amount of reads being (mis)mapped against reference targets when none are expected, for instance (visible in the PAV-spotter output figures and the IGV screenshots). Occasionally, these consist of (clipped/partially matching) reads, something that can be caused by sequencing errors, chimeric reads, errors in the reference sequence, tandem duplications, or genomic rearrangements and structural variants [69, 70].

A solution to remove any unwanted (mis-)mapped reads would be to filter more strictly upstream. However, in case of samples or targets that show poor or limited coverage – an issue that is not uncommon with target capture procedures [71] – strict filtering may not be desired. This means that, due to this potential stochasticity, merely using existing tools to check whether there are any mapped reads at all in a sample/target (i.e. checking for the presence of zero reads versus >0 reads), or building *de novo* assembled contigs per target on a sample-per-sample basis, will not be sufficient to identify PAV accurately. Our method offers an alternative solution, as PAV-spotter appears robust enough to detect PAV, even in the face of low coverage and (partially) mis-mapped and clipped reads. However, the trade-off in false positive and false negative outcomes will largely depend on the similarity thresholds and other criteria set by users, as well as on any manual, double-checks performed.

In conclusion, we show that considering the genomic position as a variable for signal displacement instead of time – which is generally the case in more classic cross-correlation applications [29] – makes it possible to identify markers of PAV/structural variation in target capture data, without needing any prior knowledge on large-scale, genomic context. Our study shows that a multidisciplinary bioengineering and biotechnological approach can help bioinformatics, and thus the fields of evolutionary and molecular research, forward when dealing with challenging research questions, datasets, and study organisms.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 802759). Sample collection was supported by the Serbian Ministry of Science, Technological Development and Innovation (grants nos. 451-03-66/2024-01/200007, 451-03-65/2024-03/200178, 451-03-66/2024-03/200178). Dr. Ana Ivanović supported breeding and collecting *Triturus* samples. Individuals of *T. ivanbureschi* and *T. macedonicus* were originally collected from natural populations: *T. ivanbureschi* from Zli Dol, Serbia with permission obtained from the Serbian Ministry of Energy, Development and Environmental Protection (permit No. 353-01-75/2014-08) and *T. macedonicus* in Ceklin, Montenegro with permission obtained from the Agency for Environmental Protection, Montenegro (permit No. UPI-328/4).The experimental procedures were approved by the Ethics Committee of the Institute for Biological Research „Siniša Stanković", University of Belgrade (decisions no. 03-03/16 and 01-1949). All experiments were performed in accordance with the Directive 2010/63/EU.

Data accessibility and Benefit-sharing statement

The raw, Illumina sequencing reads used in this study have been submitted to the NCBI Sequence Read Archive (SRA) and are publicly available through BioProject number 'PRJNA1111729' (https://www.ncbi.nlm.nih.gov/sra/PRJNA1111729). The PAV-spotter tool, including further explanation on how to use and customize the code, is available through the GitHub repository: https://github.com/Wielstra-Lab/PAVspotter. Also, certain Supplementary Materials are provided through an online Zenodo repository: https://zenodo.org/records/13991751. Figures created by PAV-spotter for each separate analysis that are not shown in the paper can be provided upon request. The same goes for all IGV screenshot images generated by the batch script.

Author contributions

MdV and CvdP designed the tool. BW, MdV and CvdP designed the experiments. MC, TV and AI collected the samples and conducted the phenotypic classification. MdV and AT performed the molecular laboratory work. CvdP wrote the PAV-spotter script (Script 2), which comprises of the signal cross-correlation code. Other scripts (Script 1, Script 3, and the IGV batch script) were written by MdV. MdV conducted the main bioinformatics, data acquisition, and data interpretation. MdV drafted the work and CvdP added mathematical details to the text and conducted p-value/threshold estimations. All authors revised the manuscript and approved of the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Annavi, G., J.E. Uli, and R. Arumugam, *A review of the application of Next Generation Sequencing (NGS) in wild terrestrial vertebrate research*. Annual Research & Review in Biology, 2019: p. 1-9.
2. Morganti, S., et al., *Next generation sequencing (NGS): a revolutionary technology in pharmacogenomics and personalized medicine in cancer*, in *Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics*, E.A.-d.l.V. Ruiz-Garcia, H., Editor. 2019, Springer Nature Switzerland: Advances in Experimental Medicine and Biology. p. 9-30.
3. Lemmon, E.M. and A.R. Lemmon, *High-throughput genomic data in systematics and phylogenetics*. Annual Review of Ecology, Evolution, and Systematics, 2013. **44**(1): p. 99-121.
4. Ekblom, R. and J. Galindo, *Applications of next generation sequencing in molecular ecology of non-model organisms*. Heredity, 2011. **107**(1): p. 1-15.
5. Wachi, N., K.W. Matsubayashi, and K. Maeto, *Application of next-generation sequencing to the study of non-model insects*. Entomological Science, 2018. **21**(1): p. 3-11.
6. Zaharias, P., et al., *Data, time and money: evaluating the best compromise for inferring molecular phylogenies of non-model animal taxa*. Molecular Phylogenetics and Evolution, 2020. **142**: p. 106660.
7. Rovelli, V., et al., *Genotyping-by-Sequencing (GBS) of large amphibian genomes: a comparative study of two non-model species endemic to Italy*. Animal Biology, 2019. **69**(3): p. 307-326.
8. Etherington, G.J., et al., *Sequencing smart: de novo sequencing and assembly approaches for a non-model mammal*. Gigascience, 2020. **9**(5).
9. Da Fonseca, R.R., et al., *Next-generation biology: sequencing and data analysis approaches for non-model organisms*. Marine Genomics, 2016. **30**: p. 3-13.
10. Andermann, T., et al., *A guide to carrying out a phylogenomic target sequence capture project*. Frontiers in Genetics, 2019. **10**: p. 1407.
11. Puritz, J.B. and K.E. Lotterhos, *Expressed exome capture sequencing: a method for cost-effective exome sequencing for all organisms*. Molecular Ecology Resources, 2018. **18**: p. 1209-1222.
12. Kaur, P. and K. Gaikwad, *From genomes to gene-omes: exome sequencing concept and applications in crop improvement*. Frontiers in Plant Science, 2017. **8**: p. 1-7.
13. Yohe, L.R., et al., *Evaluating the performance of targeted sequence capture, RNA-Seq, and degenerate-primer PCR cloning for sequencing the largest mammalian multigene family*. Molecular Ecology Resources, 2020. **20**: p. 140-153.
14. Bi, K., et al., *Transcriptome-based exon capture enables highly cost-effective comparative genomic data colection at moderate evolutionary scales*. BMC Genomics, 2012. **13**(403): p. 1471-2164.
15. Gauthier, J., et al., *A brief history of bioinformatics*. Briefings in Bioinformatics, 2019. **20**(6): p. 1981-1996.
16. Johnson, M.G., et al., *HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment*. Applications in Plant Sciences, 2016. **4**(7).
17. Yuan, H., et al., *Assexon: assembling exon using gene capture data*. Evolutionary Bioinformatics, 2019. **15**: p. 1176934319874792.
18. Andermann, T., et al., *SECAPR-a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments*. PeerJ, 2018. **6**: p. e5175.
19. Zhang, L.M., et al., *Genome-wide patterns of large-size presence/absence variants in sorghum*. Journal of Integrative Plant Biology, 2014. **56**(1): p. 24-37.
20. Wang, Y., et al., *Exploration of presence/absence variation and corresponding polymorphic markers in soybean genome*. Journal of Integrative Plant Biology, 2014. **56**(10): p. 1009-19.
21. Wold, J., et al., *Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern*. Molecular Ecology, 2021. **30**(23): p. 5949-5965.
22. Mahmoud, M., et al., *Structural variant calling: the long and the short of it*. Genome Biology, 2019. **20**(1): p. 246.
23. Marroni, F., S. Pinosio, and M. Morgante, *Structural variation and genome complexity: is dispensable really dispensable?* Current Opinion in Plant Biology, 2014. **18**: p. 31-6.
24. Gerdol, M., et al., *Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel*. Genome Biology, 2020. **21**(1): p. 275.

25. Hu, H., et al., *Amborella gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation.* New Phytologist, 2022. **233**(4): p. 1548-1555.
26. Rosa, R.D., et al., *High polymorphism in big defensin gene expression reveals presence-absence gene variability (PAV) in the oyster Crassostrea gigas.* Developmental and Comparative Immunology, 2015. **49**(2): p. 231-8.
27. Gabur, I., et al., *Gene presence-absence variation associates with quantitative Verticillium longisporum disease resistance in Brassica napus.* Scientific Reports, 2020. **10**(1): p. 4131.
28. Wellenreuther, M., et al., *Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification.* Molecular Ecology, 2019. **28**(6): p. 1203-1209.
29. Verhaegen, M. and V. Verdult, *4.3 Random Signals*, in *Filtering and system identification: a least squares approach*. . 2007, Cambridge university press. p. 100-103.
30. Gao, Z., C. Cecati, and S.X. Ding, *A survey of fault diagnosis and fault-tolerant techniques - Part I: fault diagnosis with model-based and signal-based approaches.* IEEE Transactions on Industrial Electronics, 2015. **62**(6): p. 3757-3767.
31. Jain, A.K., R.P.W. Duin, and J. Mao, *Statistical pattern recognition: A review.* IEEE Transactions on pattern analysis and machine intelligence, 2000. **22**(1): p. 4-37.
32. Hall, D.W. and M.L. Wayne, *Ohno's "peril of hemizygosity" revisited: gene loss, dosage compensation, and mutation.* Genome Biology and Evolution, 2013. **5**(1): p. 1-15.
33. Thompson, M.J. and C.D. Jiggins, *Supergenes and their role in evolution.* Heredity, 2014. **113**: p. 1-8.
34. Gutierrez-Valencia, J., et al., *The Genomic Architecture and Evolutionary Fates of Supergenes.* Genome Biology and Evolution, 2021. **13**(5).
35. Berdan, E.L., et al., *Genomic architecture of supergenes: connecting form and function.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210192.
36. Pennisi, E., *'Supergenes' drive evolution.* Science, 2017. **357**: p. 1083.
37. Schwander, T., R. Libbrecht, and L. Keller, *Supergenes and complex phenotypes.* Current Biology, 2014. **24**: p. R288-R294.
38. Llaurens, V., A. Whibley, and M. Joron, *Genetic architecture and balancing selection: the life and death of differentiated variants.* Molecular Ecology, 2017. **26**: p. 2430-2448.
39. Joron, M., et al., *A conserved supergene locus controls colour pattern diversity in Heliconius butterflies.* PLoS Biology, 2006. **4**(10): p. e303.
40. Dufresnes, C. and P.A. Crochet, *Sex chromosomes as supergenes of speciation: why amphibians defy the rules?* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210202.
41. Lamichhaney, S., et al., *Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax).* Nature Genetics, 2016. **48**(1): p. 84-8.
42. Kupper, C., et al., *A supergene determines highly divergent male reproductive morphs in the ruff.* Nature Genetics, 2016. **48**(1): p. 79-83.
43. Tuttle, E.M., et al., *Divergence and functional degradation of a sex chromosome-like supergene.* Current Biology, 2016. **26**: p. 344-350.
44. Kay, T., Q. Helleu, and L. Keller, *Iterative evolution of supergene-based social polymorphism in ants.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210196.
45. Li, J., et al., *Genetic architecture and evolution of the S locus supergene in Primula vulgaris.* Nature Plants, 2016. **2**(12): p. 16188.
46. Li, J. and C.R. Lee, *The role of gene presence-absence variations on genetic incompatibility in Asian rice.* New Phytologist, 2023. **239**(2): p. 778-791.
47. Li, C., et al., *Tight genetic linkage of genes causing hybrid necrosis and pollinator isolation between young species.* Nature Plants, 2023. **9**(3): p. 420-432.
48. Harteveld, C.L. and D.R. Higgs, *α-thalassaemia.* Orphanet Journal of Rare Diseases, 2010. **5**(13).
49. Wielstra, B., *Balanced lethal systems.* Current Biology, 2020. **30**: p. R742-R743.
50. De Visser, M.C., et al., *Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying Triturus' balanced lethal system* bioRxiv, 2024: p. 2024.10.25.620277.
51. Macgregor, H.C. and H. Horner, *Heteromorphism for chromosome 1, a requirement for normal development in crested newts.* Chromosoma, 1980. **76**: p. 111-122.
52. France, J., et al., *Genomic evidence suggests the balanced lethal system in Triturus newts originated in an instantaneous speciation event.* bioRxiv, 2024: p. 2024.10.29.620207.

53. Sessions, S.K., et al., *Cytology, embryology, and evolution of the developmental arrest syndrome in newts of the genus Triturus (Caudata: Salamandridae)*. The Journal of Experimental Zoology, 1988. **248**: p. 321-334.
54. Wallace, H., *Abortive development in the crested newt Triturus cristatus*. Development, 1987. **100**: p. 65-72.
55. Vučić, T., et al., *The reproductive success of Triturus ivanbureschi x T. macedonicus F_1 hybrid females (Amphibia: Salamandridae)*. Animals, 2022. **12**(4).
56. Vucic, T., et al., *A staging table of Balkan crested newt embryonic development to serve as a baseline in evolutionary developmental studies*. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 2024.
57. Sims, S.H., et al., *Chromosome 1 in crested and marbled newts (Triturus) - An extraordinary case of heteromorphism and independent chromosome evolution*. Chromosoma, 1984. **89**: p. 169-185.
58. De Visser, M.C., et al., *NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts)*. bioRxiv, 2024: p. 2024.10.25.620290.
59. Danecek, P., et al., *Twelve years of SAMtools and BCFtools*. Gigascience, 2021. **10**(2).
60. Wielstra, B., et al., *Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle*. Molecular Phylogenetics and Evolution, 2019. **133**: p. 120-127.
61. Meilink, W.R.M., et al., *Determining zygosity with multiplex Kompetitive Allele-Specific PCR (mxKASP) genotyping*. bioRxiv, 2024: p. 2024.10.25.620256.
62. Zavala, E.I., et al., *Quantifying and reducing cross-contamination in single- and multiplex hybridization capture of ancient DNA*. Molecular Ecology Resources, 2022. **22**(6): p. 2196-2207.
63. Kates, H.R., et al., *The effects of herbarium specimen characteristics on short-read NGS sequencing success in nearly 8000 specimens: old, degraded samples have lower DNA yields but consistent sequencing success*. Frontiers in Plant Science, 2021. **12**: p. 669064.
64. Hart, M.L., et al., *Retrieval of hundreds of nuclear loci from herbarium specimens*. Taxon, 2016. **65**(5): p. 1081-1092.
65. Pezzini, F.F., et al., *Target capture and genome skimming for plant diversity studies*. Applications in Plant Sciences, 2023. **11**(4): p. e11537.
66. Lou, R.N., et al., *A beginner's guide to low-coverage whole genome sequencing for population genomics*. Molecular Ecology, 2021. **30**(23): p. 5966-5993.
67. Jones, M.R. and J.M. Good, *Targeted capture in evolutionary and ecological genomics*. Molecular Ecology, 2016. **25**(1): p. 185-202.
68. Feng, Y., et al., *Improved molecular diagnosis by the detection of exonic deletions with target gene capture and deep sequencing*. Genetics in Medicine, 2015. **17**(2): p. 99-107.
69. Suzuki, S., et al., *ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information*. BMC Bioinformatics, 2011. **12**: p. S7.
70. Schroder, J., et al., *Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads*. Bioinformatics, 2014. **30**(8): p. 1064-1072.
71. Bragg, J.G., et al., *Exon capture phylogenomics: efficacy across scales of divergence*. Molecular Ecology Resources, 2016. **16**(5): p. 1059-68.

Close-up Triturus ivanbureschi, *Cerkezköy, Türkiye*

Chapter 6 - Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying *Triturus'* balanced lethal system

Manon de Visser[1,2], James France[1,2], Olga Paulouskaya[1,2], Thomas Brown[3,4], Michael Fahrbach[5], Chris van der Ploeg[6,7], Ben Wielstra[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, NL
2. Understanding Evolution Group, Naturalis Biodiversity Center, Leiden, NL
3. Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, DE
4. Leibniz Institute for Zoo and Wildlife Research, Berlin, DE
5. Niedernhaller Str. 8/ 2, 74653 Criesbach, DE
6. Dynamics and Control Group, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, NL
7. Integrated Vehicle Safety Group, Netherlands Organisation for Applied Scientific Research, Helmond, NL

This manuscript has been submitted to the journal ***Nature Communications*** and is currently under review. A preprinted version of this manuscript has been published at ***bioRxiv*** (2024.10.25.620277; https://doi.org/10.1101/2024.10.25.620277).

Abstract

In a balanced lethal system, half the reproductive output succumbs. *Triturus* newts are the best-known example. Their chromosome 1 comes in two distinct versions and embryos carrying the same version twice experience developmental arrest. Those possessing two different versions survive, suggesting that each version carries something uniquely vital. With target capture we obtain over 7,000 nuclear DNA markers across the genus *Triturus* and all main lineages of Salamandridae (the family to which *Triturus* belongs) to investigate the evolutionary history of *Triturus'* chromosome 1 versus other chromosomes. Dozens of genes are completely missing from either one or the other version of chromosome 1 in *Triturus*. Furthermore, the unique gene content of 1A versus 1B is remarkably similar across *Triturus* species, suggesting that the balanced lethal system evolved before *Triturus* radiated. The tree topology of chromosome 1 differs from the rest of the genome, presumably due to pervasive, ancient hybridization between *Triturus'* ancestor and other newt lineages. Our findings accentuate the complex nature of *Triturus'* chromosome 1 – the 'bloopergene' driving the evolutionarily enigmatic balanced lethal system.

Introduction

A balanced lethal system is an extreme hereditary disease that, in diploid organisms, causes the loss of fifty percent of offspring every generation [1]. A balanced lethal system is characterized by the presence of two versions of a specific chromosome that do not recombine and that each harbor one or more unique, recessive lethal alleles [1, 2]. As a consequence, individuals heterokaryotypic for this chromosome (i.e., individuals that inherited both versions), possess all genes required for survival [1, 3]. However, given the laws of Mendelian inheritance, half the offspring in each generation comprise individuals that are homokaryotypic for the chromosome, meaning they possess either one or the other version twice [4, 5]. Such individuals miss certain crucial genes, causing them to be inviable. It seems an evolutionary paradox that, despite the huge loss of fitness incurred, balanced lethal systems have evolved repeatedly across the tree of life [2, 5-9].

Given that the two chromosome versions involved in a balanced lethal system do not recombine, they should be considered supergenes: physically linked set of genes that are inherited together [10-12]. Different versions of a supergene generally encode complex, distinct phenotypes that are maintained by balancing selection when each offers a unique advantage [13, 14]. However, suppressed recombination in supergenes – often caused by structural variation such as chromosomal inversions [10] – also comes with drawbacks. Because purifying selection is weakened under reduced recombination, supergenes tend to accumulate deleterious alleles and recessive lethality may develop in homokaryotypes over time [1, 3]. For instance, this has been proposed to have caused the evolution of the recessive lethal mutations embedded in the supergenes found in fire ants [15, 16]. Moreover, supergene-bound recessive lethal genes may originate suddenly, due to a disruptive inversion breakpoint. For example, such a breakpoint mutation affects the inversion-based supergene that underlies the complex male mating strategy polymorphism in ruffs and is lethal in the homozygous state [17, 18]. Thus, a supergene might instantly qualify as, or gradually turn into, what we call a 'bloopergene': a genetic construction that can boost the rapid evolution of adaptive traits, but likewise is an evolutionary trap because of maladaptive consequences. Following this rationale, a balanced lethal system should be considered the most extreme case of a 'bloopergene'.

The best-known example of a balanced lethal system concerns 'chromosome 1 syndrome' in the salamander genus *Triturus*: the crested and marbled newts [1, 19, 20]. Over two centuries ago, Rusconi already observed that half of *Triturus* embryos perish while still inside the egg [21]. From karyotyping studies in the 1980s it became clear that the fundamental cause of this massive die-off is a balanced lethal system [20, 22]. *Triturus* newts possess two versions of chromosome 1, dubbed 1A and 1B, that lack chiasmata and do not recombine along most of the long arm [20]. Only heterokaryotypic

(1A1B) individuals survive the balanced lethal system, while homokaryotypic individuals (1A1A and 1B1B) experience developmental arrest [19, 23-25].

All *Triturus* species share chromosome 1 syndrome, indicating that this balanced lethal system originated in their common ancestor and has, remarkably, persisted for at least 24 million years [1, 26]. On the one hand, a block-like deletion structure is known to characterize the 1A- and 1B-specific regions in the *Triturus* linkage map, suggesting a rapid origin due to a chromosomal re-arrangement – a 'cytogenetic accident' [24, 27]. On the other hand, different *Triturus* species exhibit different Giemsa C-banding patterns in both 1A and 1B, suggesting that 1A and 1B have accumulated differences in genomic content between species after the balanced lethal system became fixed [25, 28]. Thus, determining the degree of any genetic divergence across *Triturus* is required to reconstruct the ancestral constitution of chromosome 1 and help us better understand the evolutionary origin of the balanced lethal system.

Supergenes are known to arise occasionally through introgressive hybridization [29, 30] and such a scenario could theoretically kickstart a balanced lethal system by bringing together two distinct versions of a chromosome, each 'pre-loaded' with private deleterious mutations, in a single population [3]. Salamandridae, the salamander family to which *Triturus* belongs, is known to have a complex evolutionary history and bears an extensive signal of introgressive hybridization [31-33]. While the 'cytogenetic accident' hypothesis would predict that 1A and 1B originated in a single ancestral *Triturus* population [27], the possibility that 1A and 1B were brought together into a single population [3] by introgressive hybridization should not be dismissed out of hand. To test these hypotheses we can compare the evolutionary history of *Triturus*' chromosome 1A and 1B to that of the other chromosomes.

Here, we conduct a genome-wide investigation using target capture data of 148 samples from 16 genera within the Salamandridae family (covering 26 species in total), with the aim of unraveling the evolution of *Triturus*' peculiar chromosome 1. Firstly, we identify presence/absence variation associated with the balanced lethal system in *Triturus* by comparing 1A- and 1B-linked markers across ten different *Triturus* species to explore how the gene content varies. Secondly, we investigate the evolutionary history of *Triturus'* chromosome 1 versus that of the rest of the *Triturus* genome by conducting phylogenomic analyses including all main Salamandridae lineages. We interpret our results in the light of existing and new theories regarding the origin of balanced lethal systems.

Methods

Sampling scheme and phenotypic + genotypic classification

To investigate the intergeneric variation of *Triturus*' chromosome 1A and 1B, we obtain embryos from captive populations of all *Triturus* species listed in Table S1 [34]. We include both the Balkan and Italian clade of *T. carnifex*, which are genetically distinct and presumably represent distinct species [35]. We do not include *T. rudolfi* [36], closely related to *T. pygmaeus*, as it was only described after our sampling. Individuals were (pure)bred for an unknown number of generations in captivity, but the original founders of all breeding lines are known to derive from localities away from any (potential) hybrid zones [37], ensuring interspecific admixture is limited. Husbandry and breeding practices are described in [38].

The two types of homokaryotypic *Triturus* embryos (1A1A and 1B1B) express different phenotypes upon developmental arrest: the so-called "slim-tailed" (1A1A) embryos appear morphologically similar to healthy heterokaryotypic (1A1B) embryos, whereas the so-called "fat-tailed" embryos (1B1B) show more severe malformations [24, 25, 39]. During the sampling of *Triturus* embryos, embryos are thus classified as being healthy when they developed beyond the critical late tailbud stage, or as being diseased (in one, or the other, phenotypic group) when they experienced developmental arrest during this critical stage. In addition, we confirm this embryonic classification by genotyping samples using a 1A-linked marker (PLEKHM1), a 1B-linked marker (NAGLU), and a control marker from elsewhere in the genome (CDK) with multiplex PCR [40, 41]. In total, we include 90 *Triturus* embryo samples: three individuals per 1A1B, 1A1A, and 1B1B genotype, for ten species/populations.

For building a *Triturus* species tree, we use previously published data of 27 adult *Triturus* newts [42, 43] and add three *T. carnifex* samples [35] to account for presumed species-level divergence within this taxon (Table S1). Furthermore, to determine the phylogenetic relationships of 1A and 1B compared to the rest of the *Triturus* genome, we include other newts from the Salamandridae family in our study: we use three samples for closely related lineages (the modern European newts, comprising of the genera *Lissotriton, Ommatotriton, Ichthyosaura, Neurergus* and *Calotriton*) and at least one sample as representative for more distantly related lineages [33, 44]. This means that, next to the *Triturus* samples, we use 28 Salamandridae samples from 15 different genera (Table S1).

Laboratory procedures and pre-processing of sequence data

To obtain genomic data we use "NewtCap": a target enrichment by sequence capture workflow [44]. In brief, we apply a standard salt-based method for extracting DNA using

the Promega Wizard™ Genomic DNA Purification kit (Promega, Madison, WI, USA), followed by library preparation using the NEBNext Ultra™ II FS DNA Library Preparation Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Then, we perform target enrichment using a MyBaits-II kit using sequence-specific, *Triturus*-based RNA probes [Arbor Bioscience, Ann Arbor, MI, USA, product Ref# 170210-32; 42, 44] and Illumina sequencing of the enriched targets (outsourced to Baseclear B.V., Leiden, the Netherlands). An upstream bioinformatics pipeline is used to check the quality of the reads and to perform clean-up and mapping upstream [44]. Downstream analyses are described below.

Oxford Nanopore sequencing

To gather as much information about the (likely) genomic position of markers within the DNA of *Triturus*, we perform whole genome sequencing, followed by *de novo* assembly. We use liver tissue from a *T. ivanbureshi* individual, flash frozen in liquid nitrogen and stored at -80 °C prior to whole genome sequencing. High molecular weight DNA extraction and DNA sequencing using Nanopore is performed by Future Genomics Technologies. DNA is prepared using the SQK-LSK110 ligation library kit. FLO-MIN106 and FLO-PRO002 (R9.4.1) flow cells are used for sequencing on MinION and PromethION platforms. For basecalling, Guppy v.5.0.17 [45] is used with the high-accuracy model. From 14 flowcells, we obtain 91,494,016 reads, with an N50 ranging from 14,7 kb to 27,2 kb per flow cell. Using Porechop [46] we trim adaptor sequences off of the raw reads.

Genome assembly, scaffolding and alignment

We assemble the genome of *Triturus de novo* using the Oxford Nanopore Technology reads and scaffold the rather fragmentary output using the *Pleurodeles waltl* genome (56) as a reference to get a chromosome-level assembly. Although *P. waltl* does not suffer from a balanced lethal system, it is the most closely related genus with a published reference genome available and synteny with *Triturus* is high [27], so it allows us to estimate the order of genes on *Triturus'* chromosome 1.

We run Shasta v.0.10.0 [47] with arguments: "--Reads.minReadLength 5000" and "--config Nanopore-May2022", using reads > 5 kb, followed by removal of any retained haplotigs using purge-dups v.1.2.6 [48] to produce a draft *Triturus* genome assembly of 21 Gb in size. The assembly contains ~65k contigs, with an N50 of 1.27 Mb. For scaffolding, we run RagTag [49] with the *P. waltl* genome as input reference. This eventually scaffolds ~17.6Gb of the 21Gb contigs assembly into the 12 separate chromosomes. We then use BLAST+ v.2.2.3 1 [50] to align the genes from the target capture set against our scaffolded genome assembly and *P. waltl* reference genome with default parameters. Only the genes that have one hit in each of the genomes are considered for further analysis.

Presence/absence variation analysis with PAV-spotter

We use "PAV-spotter" [41] to search for presence/absence variation in the target capture data of three diseased fat-tailed (1B1B), three diseased slim-tailed (1A1A), and three healthy/control (1A1B) embryos, for each *Triturus* species/population independently. PAV-spotter uses BAM file information to calculate signal-cross-correlation values that are based on the resemblance of the distribution of reads mapped against reference sequences for a given target. We use the following rationale: an absence of reads for a target across 1B1B samples is indicative of a 1A-linked marker in case reads are mapped for the same target across 1A1A and 1A1B samples, and likewise an absence of reads for a certain target across 1A1A samples denotes a 1B-linked marker when reads are mapped for the same target across 1B1B and 1A1B samples.

We prepare the read depth data for PAV-spotter and we apply the tool using default settings [41]. We use three samples per phenotypic category to correct for any stochasticity and variation in capture-rates between samples. To identify chromosome 1-linked presence/absence variation, we used a 80% dissimilarity threshold [41]. We confirm the presence/absence variation patterns by visually inspecting the read content in the BAM files for targets that are identified as either 1A- or 1B-linked markers using IGV [Integrative Genomics Viewer; 51] and we again do this in the same way as described in [41].

Reconstructing the ancestral constitution of chromosome 1

After obtaining the list of candidates for 1A-linked and 1B-linked markers, we cross-compare the results for the different *Triturus* species. To determine if targets are true positives, we cross-check the data with 1) a *Triturus* linkage map [27], 2) a *Lissotriton* linkage map [27], 3) the *Pleurodeles* genome assembly [52], and 4) the *Triturus* RagTag assembly (which is based on the *Pleurodeles* genome). In case targets do not show up in expected regions within these comparative genomic datasets, they are excluded as false positives (approximately 70% of potential 1A- and 1B-linked markers concerns singletons, meaning they are only discovered in one species).

As synteny across newt genera is high, but not absolute [27], we follow a rationale in the order of cross-comparisons that is based on the relatedness to *Triturus*, where data of more closely related genera overrule the data of more distantly related genera. If a potential 1A- or 1B-linked target is found in the region of the *Triturus* linkage map that corresponds to the region of interest on chromosome 1 (between ~0–50 cM on linkage group 1), we consider it a true positive. Next, for any remaining target that is not present in the *Triturus* linkage map, we consider it a true positive if it is present on the

relevant region of the *Lissotriton* linkage map (between ~50–100 cM on linkage group 4). For any remaining target at this point, we perform a cross-check with the best blast hits to the less closely related genus *Pleurodeles*. In case the target is recovered with a best blast hit on *Pleurodeles'* scaffold 6 between ~0–500 Mb (which corresponds to the region of interest in *Triturus*), we include it. Finally, for any remaining target we cross-check against our *Triturus* RagTag assembly. In case we observe that a target is on scaffold 6 of the RagTag assembly between ~0–500 Mb (again, the region of interest in *Triturus*), we include it. Any remaining targets are excluded.

Building the Triturus phylogeny

To help visualize potential phylogenetic signal in the presence/absence variation matrix, we build a *Triturus* phylogeny. We exclude the entirety of chromosome 1 by restricting the input for the analysis to the 3,728 targets that were placed on linkage groups 2-12 in the *Triturus* linkage map [27]. To extract only high quality SNPs, we apply quality filtering by removing heterozygote excess, as well as by extracting SNPs and discarding INDELs, applying stringent quality filtering options, and removing sites with more than 50% missing data [as described in; 44]. Next, we remove invariant sites using an ascertainment bias correction Python script (https://github.com/btmartin721/raxml_ascbias) and convert the filtered VCF to PHYLIP format using the vcf2phylip script, v.2.8 [53] to serve as input for RAxML. Then, we run RAxML v.8.2.12 with 100 rapid bootstrap replicates under the ASC_GTRGAMMA model and Lewis ascertainment correction to perform a concatenation analysis and obtain the best-scoring maximum likelihood tree [54], following the methods of [44]. This phylogeny (Fig. 1 and Fig. S35) was based on in total 19,827 SNPs. We visualized the tree using FigTree v.1.4.4 and enhanced the coloured version in Figure 1 using iTOL v.6 [55].

Estimating Hardy-Weinberg Equilibrium deviations

To search for genes that are potentially still present on both chromosome 1A and 1B across (most of the) *Triturus* species, we search for targets that deviate from Hardy-Weinberg Equilibrium and display excessive heterozygosity explicitly in 1A1B embryos (and not in 1A1A or 1B1B embryos). We create three multi-sample VCF files for *Triturus*; one with the 30 1A1B embryos, one with the 30 1A1A embryos, and one with the 30 1B1B embryos. Then, we use BCFtools v.1.18 [56, 57] to estimate Hardy-Weinberg and related statistics for every site in each of the VCF file. To take into account potential gene losses, or stochasticity in the data, we allow for a maximum of three individuals to have a missing genotype at any particular site per each VCF file, and we filter as such using the *--max-missing* function of VCFtools [58].

From the data we extract the sites for which heterozygote excess was discovered in the 1A1B set (with a p-value below 0.05), but neither in the 1A1A, nor in the 1B1B set (i.e., the p-value was above 0.05 in both). As before (see "Presence/absence variation analysis with PAV-spotter"), we check whether the discovered targets fall in the (vicinity of the) heteromorphic *Triturus* chromosome 1 region. Finally, we again perform a visual inspection in IGV [51] to determine if heterozygous SNPs are restricted to 1A1B embryos. Based on a p-value of 0.05, we discover only six targets that show significant heterozygote excess in explicitly 1A1B embryos, pass our IGV inspection, and appear in or near the region(s) of interest (Table S1).

Phylogenetic analyses of Salamandridae

To investigate if the topology of chromosome 1 differs from that of the rest of the genome, we build a species tree of Salamandridae, following the same phylogenetic approaches as described above. We perform separate RAxML analyses based on SNPs from chromosome-level subsets of targets mapped to different *Triturus* linkage groups: "Chr1" (all 390 targets on *Triturus* linkage group 1, including 1A- and 1B-linked markers), "Chr1*" (a subset of 329 targets on *Triturus'* recombining part of linkage group 1, thus excluding 1A- or 1B-linked markers), "1A" (the subset of 28 markers that was 1A-linked for all *Triturus* species analyzed), "1B" (the subset of 35 markers that was 1B-linked for all *Triturus* species analyzed), "Chr2-12" (all 3,728 targets on *Triturus* linkage groups 2-12), and "Chr2" – "Chr12" (subsets of targets for each individual linkage group; see Table S3). We time-calibrate the best-scoring ML trees discovered by RAxML using TreePL [59] using to strict calibration points: 67.49 MYA, for the split between the "New World Newts" and "Old World Newts" and 24 MYA for the most recent common ancestor of *Triturus* [26]. The trees are visualized in FigTree v.1.4.4 and shown in Figs. S1-S34.

Calculating D- and f-statistics

To test the hypothesis that chromosomes 1A and 1B originated in different evolutionary newt lineages and later ended up in a single population through introgressive hybridization, we calculate Patterson's D, also known as the 'ABBA-BABA statistic' [60] and related statistics such as the admixture fraction f (also known as the 'f_4-ratio) using Dsuite [60, 61]. We focus on the Modern European Newts, which comprise *Triturus*, *Lissotriton*, the NIO clade (*Neurergus*, *Ichthyosaura* & *Ommatotriton*) and *Calotriton* [33]. Calculations are based on the total dataset and on chromosome-level datasets (Table S3).

We conduct the Dsuite analyses for the three different topologies discovered with the phylogenetic analyses earlier: the general species tree based on linkage groups 2-12, with *Lissotriton* being the closest relative to *Triturus*, followed by the NIO clade, then followed by *Calotriton* – named 'TOP1'), the alternate species tree as inferred with targets from chromosome 1 (with the NIO clade being closer related to *Triturus* than sister lineage

Lissotriton – named 'TOP2'), and the deviant species tree as inferred with targets from only 1B (with both *Calotriton* and the NIO clade being more closely related to *Triturus* than sister lineage *Lissotriton* – named 'TOP3'). The topologies are provided to Dsuite in Newick format. We use *Euproctus* as the outgroup. The Dtrios function of Dsuite is used to calculate the D- and f_4 admixture ratio statistics for all potential trio combinations. We apply Dsuite's Fbranch function on trios with significantly elevated D-values to calculate f-branch statistics, which are related to the f4-ratio. Finally, we construct heatmaps using Dsuite as well to visualize the results (see also Table S3 and Figs. S36 – S86).

Determining the number of gene flow events and their direction

The software Treemix allows us to infer maximum likelihood trees and details on admixture events from our target capture data [62, 63]. As input we use the cleaned-up msVCF files that include the Modern European Newt samples with high quality SNPs. We perform more filtering to ensure files have no missing data using VCFtools v.0.1.16 [58] and we remove any sites that are in linkage disequilibrium using PLINK v.1.9 [64]. We then run the linkage disequilibrium-pruned data through Treemix, using *Calotriton* as an outgroup with 500 bootstraps, for one through five possible migration edges (m), and for an iteration of ten analyses. Then, we use the R package "OptM" [65] to determine the optimal value of m using the Treemix output of all iterations. Given that the variation in the standard deviations of the input data of our subsetted datasets appears too low, OptM is not able to retrieve an optimal value for m for smaller subsets of the data (i.e., when including only markers that belong to a certain linkage group), so we decide to only include the overall target capture dataset. We visualize results using (adjusted versions of) the plotting_funcs.R script of Treemix and present the Evanno and Linear method outcomes by using the plot_optM function of OptM (Figs. S87 and S88).

Results

Consistent deletions on 1A and 1B across *Triturus*

To test the hypothesis that the 1A- and 1B-specific regions of different *Triturus* species vary in genomic content, we use target enrichment by sequence capture to aim to obtain 7,139 nuclear DNA markers [42, 44] from 1A1A, 1A1B and 1B1B embryos belonging to ten *Triturus* species (n=3 for each genotype, for each species, as listed in Table S1). In healthy (1A1B) embryos, 6,785 genes are recovered. We explore which of these genes are missing in either 1A1A embryos (and therefore presumably absent from chromosome 1A) or 1B1B embryos (and therefore potentially absent from chromosome 1B) for each

Triturus species. To strengthen our confidence that genes are truly absent from 1A or 1B, we determine their position on linkage maps of *Triturus* and its sister genus *Lissotriton;* [the smooth newts; 27], the chromosome-scale assembly of the distantly related *Pleurodeles;* [the sharp-ribbed newt; 52], and an Oxford Nanopore sequencing-based *Triturus* assembly, scaffolded against the *Pleurodeles* genome, with the understanding that synteny in newts is high [27]. We find that 29 genes are consistently absent *Triturus*-wide in diseased 1A1A embryos (but present in 1A1B and 1B1B embryos; Fig.1 and Table S2) and another 35 genes are consistently absent across *Triturus* in diseased 1B1B embryos (but present in 1A1B and 1A1A embryos; Fig.1 and Table S2).

The 1A and 1B chromosomes are remarkably similar in the different *Triturus* species in terms of gene content: variation in gene-deletions is low between species, and there is no apparent phylogenetic signal for the few gene-deletions that do deviate between species (Fig. 1). This indicates that deletions mainly happened before the radiation of *Triturus* and that only a few genes were lost afterwards. Also, the gene order along the chromosome-scale genome assembly of *Pleurodeles waltl* [52], suggests that chromosomes 1A and 1B are characterized by unique, big deletions of DNA stretches that contain dozens of genes and together span ~600 Mbp (in *Pleurodeles*). The genes deleted from 1A- and 1B generally form two consecutive clusters, followed by a small 'mixed' block consisting of both presence/absence variation markers, as well as markers with distinct copies on A and B (Fig.1 and Table S2). This 'mixed' block suggests the presence of a relatively small, evolutionary stratum present on both 1A and 1B, adjacent to the 1A-block in the estimated, ancestral constitution of chromosome 1.

Given the conserved synteny across different, distantly related newt genera [27], and given our observation of little variation in 1A- and 1B-linked markers across the entire genus *Triturus*, we can now further explore the evolutionary history of chromosomes 1A and 1B compared to the rest of the *Triturus* genome.

Discordant evolutionary history for *Triturus'* chromosome 1

To test the hypothesis that chromosomes 1A and 1B originated in different evolutionary lineages of newt and were later brought together in a single population through introgressive hybridization, we sequence our c. 7k nuclear DNA markers for all main evolutionary lineages in the Salamandridae family (Table S1). We then build phylogenetic trees for markers from all twelve linkage groups (presumed to correspond to the twelve chromosomes) of *Triturus* separately [27], and we also build trees for only the recombining section of chromosome 1 (i.e. predominantly the short arm), the 1A-linked markers, and the 1B-linked markers, to explore whether either the 1A- or the 1B-linked markers display an aberrant evolutionary history.

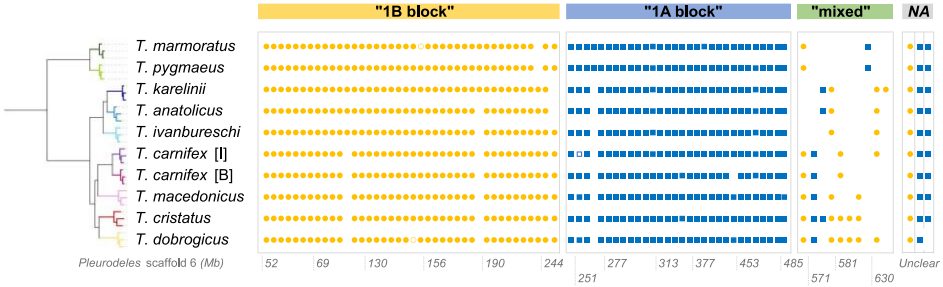


Fig 1. *Presence/absence variation matrix showing variation in 1A- and 1B-linked markers across the genus* Triturus*. For each* Triturus *species, three individuals with the 1A1A genotype, three with the 1A1B genotype, and three with the 1B1B genotype, were compared. Each column in the matrix represents a recovered 1A-linked (blue square) or 1B-linked marker (orange circle). A gap indicates no presence/absence was detected and an open symbol reflects uncertainty. The matrix is ordered based on the position of the markers in the* Pleurodeles *genome (scaffold 6; location in Mbp indicated at the bottom). From left to right there is a distinct 1B-block containing 40 markers, followed by a distinct 1A-block containing 30 markers, followed by a relatively small 'mixed' block with three 1A- and seven 1B-linked markers. Three additional markers could not be mapped in* Pleurodeles *('Unclear'). A total of 35 1B-linked and 28 1A-linked markers are consistently found across all* Triturus *species. The* Triturus *phylogeny is based on 19,827 SNPs from 3,728 targets that are not linked to chromosome 1 (see also Fig. S35). [I] = Italy clade, [B] = Balkan clade.*

We observe phylogenetic discordance in the 'modern European newts', a group of newts composed of four main lineages: the genera *Triturus*, *Lissotriton* and *Calotriton*, as well as the 'NIO clade', which comprises the genera *Neurergus*, *Ichthyosaura* and *Ommatotriton* [33]. The topology for the entire chromosome 1, rather than only the 1A-specific or 1B-specific regions, deviates from that of the rest of the genome (Fig. 2 and Figs. S1-S18). Individual phylogenies for linkage groups 2 through 12 suggest that *Lissotriton* is the sister lineage of *Triturus* and that, together, these two genera comprise the sister lineage of the NIO clade. On the other hand, both the recombining section of chromosome 1, as well as the separate 1A- and 1B-linked markers, suggest that the phylogenetic positions of the NIO clade and *Lissotriton* are switched, with the NIO clade recovered as the sister lineage of *Triturus* (although for 1B-linked markers the support for *Triturus* + the NIO clade is relatively low). The 1B-linked markers show further deviation still: not *Lissotriton,* but *Calotriton* (more distantly related for all the other linkage groups) is suggested to be sister to the *Triturus* + the NIO clade, albeit with very low support (Fig. 2). Lastly, molecular dating suggests that the radiation of all newt clades involved (thus *Triturus*, *Lissotriton*, the

NIO clade, and *Calotriton*, together known as the 'modern European newts') occurred in the distant past and in a brief time window (c. 38-42 million years ago; Figs. S18-S34). Given this topological discordance, we investigate the genomic pattern of introgression in more detail.

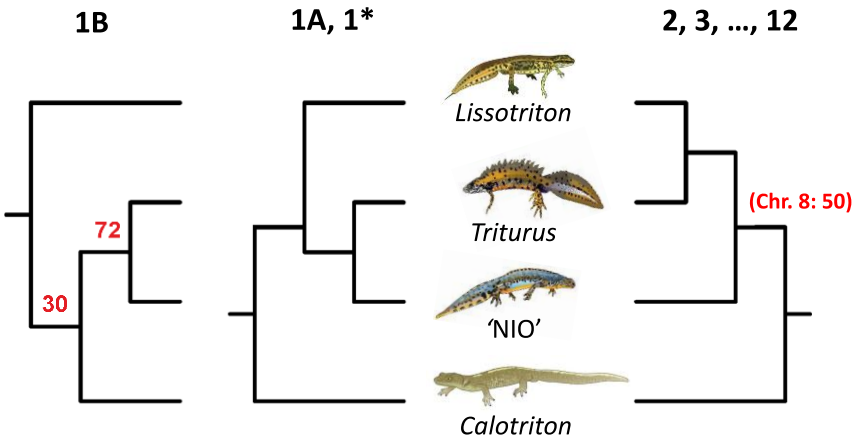


Fig 2. *Topologies for the four clades that comprise the 'modern European newts' for different parts of the genome. The topology on the far right (2-12) shows the result for targets of chromosomes 2 through 12 combined, as well as for each of the 12 linkage groups separately. The two topologies on the right show the result for the recombining region of chromosome 1 (1*), for chromosomes 1A (same as 1*), and for chromosome 1B – together comprising linkage group 1. Red numbers indicate three cases where the bootstrap support was below 80 (see Figs. S1-S18). The NIO clade is a consistently recovered monophyletic group that contains the three genera* Neurergus, Ichthyosaura *and* Ommatotriton.

A complex history of pervasive introgression in modern European newts

The discordant phylogenetic position of *Triturus* within the modern European newts, based on chromosome 1 versus the rest of the genome, implicates a history of introgression. We test for deviations from a strictly bifurcating evolutionary history in order to identify potential key introgression events, again for markers positioned on the different *Triturus* linkage groups, including 1A, 1B and the recombining section of chromosome 1. Our analyses show that a signal of introgression involving the NIO clade affects not only *Triturus'* chromosome 1, but most other linkage groups as well (Fig. 3). To a lesser extent, introgression involving *Calotriton* is also implicated, which is in line with the observation that mtDNA suggests this taxon to be the sister lineage of *Triturus* [33]. Overall, the dominant signal of introgressive hybridization suggests introgression from the ancestor of the NIO clade into the ancestor of the *Triturus* lineage (Fig. 4).

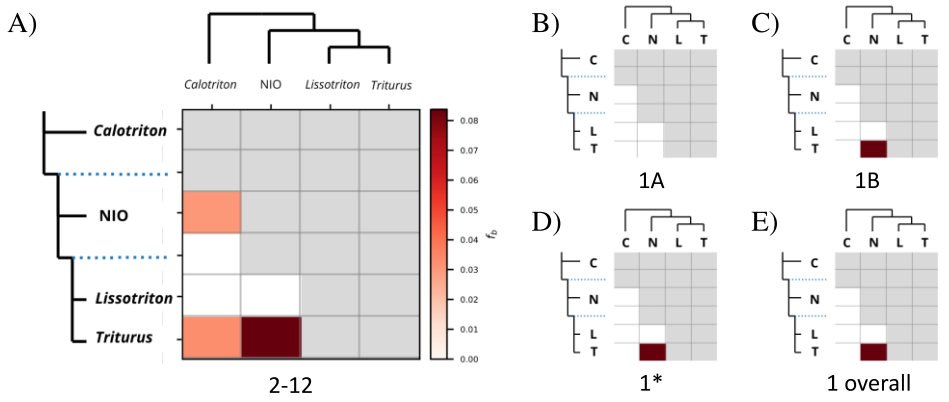


Fig 3. *F-branch (f_b) heatmaps for different subsets of markers/linkage group(s), showing signals of pervasive introgression in the past among 'modern European newts'. Colour shading reflects the intensity of excess allele-sharing (indicative of introgression) between tree branches on the y-and x-axes. Grey boxes reflect it is not possible to calculate f-branch statistics following the tree topology. Blue, dotted lines correspond to the internal tree branches (e.g. the line above* Lissotriton *and* Triturus *represents their common ancestor). (A) Heatmap based on markers positioned on* Triturus *linkage groups 2 through 12 (n=3,728 targets). (B, C, D, E) Summarized heatmaps for, respectively; 1A-linked markers only (n=28 targets), 1B-linked markers only (n=35 targets), markers positioned on the recombining region of chromosome 1 (1*, n=329 targets), and all markers found on chromosome 1 (n=390 targets). A noticeable, strong signal for introgression between* Triturus *and the NIO clade is visible, except in the analysis including 1A-linked markers only (which notably has the lowest sample size). Abbreviations in figure B-E: T =* Triturus*, L =* Lissotriton*, N = the NIO clade (a consistently recovered monophyletic group that contains the three genera* Neurergus*,* Ichthyosaura *and* Ommatotriton*), and C =* Calotriton*. See Table S3 and Figs. S36 – S86 for all heatmaps produced.*

Discussion

The chromosome underlying the balanced lethal system in *Triturus* has a remarkable evolutionary history. Gene-content-wise, little has changed in 1A and 1B since the balanced lethal system became established in the *Triturus* ancestor. This is surprising, because natural selection would be expected to be less efficient at preventing the further shedding of genes since the system became fixed, as homokaryotypes from that point on would have been destined to die anyway [1, 3, 66].

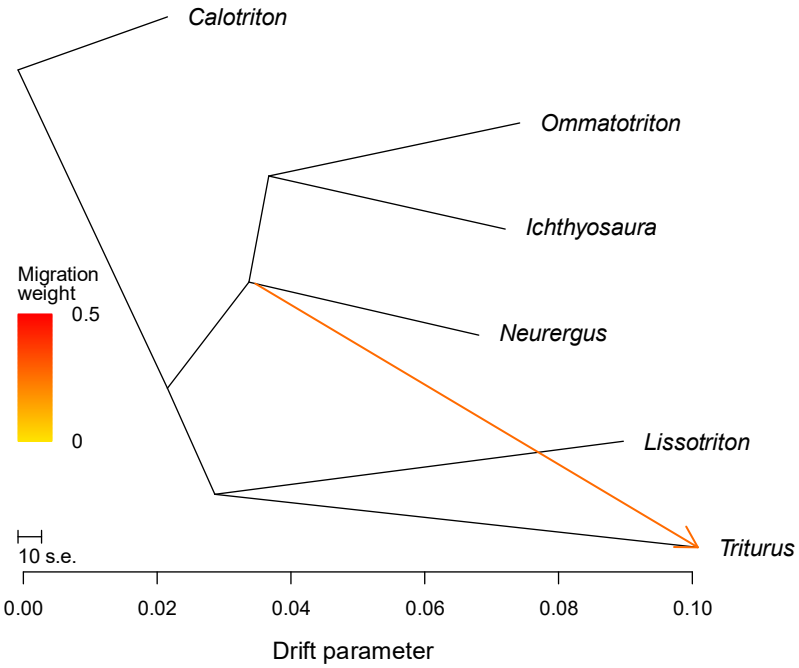


Fig 4. *TreeMix admixture graph for the 'modern European newts' with one migration edge, based on the overall target capture dataset. The arrow suggests that introgression occurred between the ancestors of the NIO clade and the Triturus genus and shows the inferred direction of the migration event. The colour of the arrow indicates the migration weight. The axis at the bottom shows the genetic drift measured in the populations.*

We see that the 1A- and 1B-specific regions together span ~600 Mbp in the genome of *Pleurodeles* [52], which is considerably smaller than would be expected based on the physical size inferred in karyotyping studies of *Triturus*, estimated to be at least 1.2 Gb [25] – over a third of the size of the human genome. The affected region includes a small, evolutionary stratum that suggests restriction of recombination extends beyond the core 'bloopergene', allowing further divergence to have accumulated between 1A and 1B (although the pattern is notably minor). The variation in Giemsa C-banding patterns that is observed between different lineages of the genus presumably reflects the independent accumulation of non-coding DNA [25, 28]. This is not unexpected, given that salamander genomes are dominated by transposable elements, which tend to accumulate in non-recombining regions with weak, or absent, selection pressures [67-69]. However, even though the non-recombining long arms of chromosome 1 – comprising 1A and 1B – are physically large, the gene density is evidently low.

The near-identical gene content of 1A and 1B across the genus *Triturus* (in combination with the lack of a phylogenetic signal for the minor variation between species that is observed), supports that the 'bloopergene' driving the *Triturus* balanced lethal system originated *before* the onset of the radiation of *Triturus* occurred. This fits the hypothesis that chromosomes 1A and 1B evolved instantaneously due to a 'cytogenetic accident' [24, 27]. But does that preclude that introgressive hybridization played a role in the origin of the balanced lethal system?

Our phylogenomic approach reveals a divergent topology for chromosome 1 compared to the rest of the *Triturus* genome. The entire *Triturus* genome bears the signature of introgression derived from the NIO clade. However, introgressive hybridization seems to have affected chromosome 1 in particular, to the extent that a sister relationship between *Triturus* and the NIO clade is supported in our phylogenetic analysis (rather than *Lissotriton* being the sister lineage of *Triturus* as is the case for the rest of the genome). Yet, the scenario that chromosome 1 reflects the actual evolutionary relationships and was relatively resistant to introgression from *Lissotriton* compared to the rest of the genome can perhaps not be dismissed out of hand, given that including more molecular markers does not strictly lead to a more accurate species tree [70-72].

Our findings do not fit the hypothesis that introgression brought together supergenes that evolved in two different ancestral populations [3]. The confounding topology was observed not only for 1A-specific regions or for 1B-specific regions of chromosome 1, but for both of them – as well as for the rest of chromosome 1. This further supports that the balanced lethal system arose strictly within the ancestral *Triturus* population – in line with the hypothesis that chromosomes 1A and 1B originated in a cytogenetic accident [24, 27]. Still, the aberrant evolution of chromosome 1 in *Triturus* evidently did involve introgressive hybridization.

In conclusion, we reconstruct the ancestral constitution of *Triturus'* chromosome 1 and provide new insights into the evolutionary mechanisms that twisted the *Triturus* genome in the distant past. The 'bloopergene' that drives the balanced lethal system evolved over 24 million years ago within the ancestor of *Triturus*, against a backdrop of pervasive, introgressive hybridization, involving the other lineages that gave rise to the modern European newts. By unveiling the intricate genomic architecture and multifaceted history of the *Triturus* balanced lethal system, our results – paradoxically, much like the phenomenon itself – both enhance and challenge our understanding of how supergenes and 'bloopergenes' exactly originate and evolve in nature.

Permit section

Triturus embryos were obtained by M.F. from his personal breeding colony. Housing and breeding salamanders as a private individual are not considered to require a license (BGBl. I S. 3125, 3126, 3750). Housing and breeding protocols comply to EU directive standards (EU directive annex III, section B, Table 9.1) and were reviewed by the Animal Welfare Body Leiden. As per EU legislation regarding the protection of animals used for scientific purposes (EU directive no. 2010/63/EU), sacrificing embryos that are not feeding independently does not qualify as an animal experiment. Sacrificing a *Triturus* individual for the Oxford Nanopore sequencing was approved by the Centrale Commissie Dierproeven (CCD; decision no. AVD1060020198065) after advice of the independent Dierexperimentencommissie Leiden (DEC Leiden).

Acknowledgements

Funding was provided by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 802759). T.B. was supported by DFG (INST 269/768-1). This work was performed using the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. We are grateful for the feedback provided to us by Professor Roger K. Butlin from the University of Sheffield, and we thank Erik-Jan Bosch for creating the scientific illustrations of the salamanders under the open content License of Naturalis Biodiversity Center (© CC BY-NC-ND 4.0).

Author contributions

M.d.V., J.F. and B.W. conceived and designed the research. M.F. collected newt embryos. M.d.V. and J.F. performed the lab-work and pre-processing of the data. M.d.V. conducted the downstream analyses, with the assistance of O.P., C.v.d.P., T.B., and J.F. M.d.V., J.F. and B.W. wrote the draft version of the paper, and all authors contributed to revising it.

Competing interests

The authors declare no competing interests.

Data availability

Illumina sequencing reads are available through BioProject (PRJNA1173497, PRJNA1171613, and PRJNA498336). Oxford Nanopore Technology sequencing reads will also be available (*source; TBD*). The main scripts used for downstream analyses (Dsuite and Treemix) are provided through GitHub (https://github.com/Wielstra-Lab/Triturus_chr1_bloopergenes). Supplementary Materials are accessible via Zenodo (https://zenodo.org/records/13991240).

References

1. Wielstra, B., *Balanced lethal systems.* Current Biology, 2020. **30**: p. R742-R743.
2. Dawson, P.S.A., *A balanced lethal system in the flour beetle Tribolium castaneum.* Heredity, 1967. **22**: p. 435-438.
3. Berdan, E.L., et al., *Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210199.
4. Mendel, G., *Versuche uber pflanzen-hybriden.* Vortlegt in den Sitzungen, 1865.
5. Dobzhansky, T. and O. Pavlovsky, *An Extreme Case of Heterosis in a Central American Population of Drosophila tropicalis.* Proceedings of the National Academy of Sciences, 1955. **41**: p. 289-295.
6. James, S.H., *Complex hybridity in Isotoma petraea. I. The occurrence of interchange heterozygosity, autogamy and a balanced lethal system.* Heredity, 1965. **20**: p. 341-353.
7. Steiner, E., *New aspects of the balanced lethal mechanism in Oenothera.* Genetics, 1956. **41**: p. 486-500.
8. Lin, Y.J., *Chromosome distribution and catenation in Rhoeo spathacea var. concolor.* Chromosoma, 1979. **71**: p. 109-127.
9. Thien, L.B., *Chromosome translocations in Gayophytum (Onagraceae).* Evolution, 1969. **23**: p. 456-465.
10. Thompson, M.J. and C.D. Jiggins, *Supergenes and their role in evolution.* Heredity, 2014. **113**: p. 1-8.
11. Schwander, T., R. Libbrecht, and L. Keller, *Supergenes and complex phenotypes.* Current Biology, 2014. **24**: p. R288-R294.
12. Hill, W.G. and A. Robertson, *The effect of linkage on limits to artificial selection.* Genetics Research, 1966. **8**: p. 269-294.
13. Black, D. and D.M. Shuker, *Supergenes.* Current Biology, 2019. **29**(13): p. R615-R617.
14. Llaurens, V., A. Whibley, and M. Joron, *Genetic architecture and balancing selection: the life and death of differentiated variants.* Molecular Ecology, 2017. **26**: p. 2430-2448.
15. Kay, T., Q. Helleu, and L. Keller, *Iterative evolution of supergene-based social polymorphism in ants.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210196.
16. Wang, J., et al., *A Y-like social chromosome causes alternative colony organization in fire ants.* Nature, 2013. **493**(7434): p. 664-8.
17. Kupper, C., et al., *A supergene determines highly divergent male reproductive morphs in the ruff.* Nature Genetics, 2016. **48**(1): p. 79-83.
18. Hill, J., et al., *Low mutation load in a supergene underpinning alternative male mating strategies in ruff (Calidris pugnax).* Molecular Biology and Evolution, 2023. **40**(12).
19. Wallace, H., *The balanced lethal system of crested newts.* Heredity, 1994. **73**: p. 41-46.
20. Macgregor, H.C. and H. Horner, *Heteromorphism for chromosome 1, a requirement for normal development in crested newts.* Chromosoma, 1980. **76**: p. 111-122.
21. Rusconi, M., *Amours des salamandres aquatiques: et developpement du tetard de ces salamandres depuis l'oeuf jusqu'a l'animal parfait*. 1821, Milano: Giusti.
22. Horner, H.A. and H.C. Macgregor, *Normal development in newts Triturus and its arrest as a consequence of an unusual chromosomal situation.* Journal of Herpetology, 1985. **19**(2): p. 261-270.
23. D'Amen, M., L. Vignoli, and M.A. Bologna, *The normal development and the chromosome No. 1 syndrome in Triturus carnifex carnifex (Caudata, Salamandridae).* Italian Journal of Zoology, 2006. **73**: p. 325-333.
24. Sessions, S.K., et al., *Cytology, embryology, and evolution of the developmental arrest syndrome in newts of the genus Triturus (Caudata: Salamandridae).* The Journal of Experimental Zoology, 1988. **248**: p. 321-334.
25. Sims, S.H., et al., *Chromosome 1 in crested and marbled newts (Triturus) - An extraordinary case of heteromorphism and independent chromosome evolution.* Chromosoma, 1984. **89**: p. 169-185.
26. Steinfartz, S., et al., *A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary atterns of the Salamandridae with emphasis on Triturus newts.* Journal of experimental zoology. Part B, Molecular and developmental evolution, 2007. **308B**: p. 139-162.

27. France, J., et al., *Genomic evidence suggests the balanced lethal system in Triturus newts originated in an instantaneous speciation event.* bioRxiv, 2024: p. 2024.10.29.620207.
28. Callan, H.G. and L. Lloyd, *Lampbrush chromosomes of crested newts Triturus cristatus (Laurenti)*. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1960. **243**(702): p. 135-219.
29. Jay, P., et al., *Supergene evolution triggered by the introgression of a chromosomal inversion.* Current Biology, 2018. **28**(11): p. 1839-1845.
30. Stolle, E., et al., *Recurring adaptive introgression of a supergene variant that determines social organization.* Nature Communications, 2022. **13**(1): p. 1180.
31. Veith, M., et al., *The changing views on the evolutionary relationships of extant Salamandridae (Amphibia: Urodela)*. PLoS ONE, 2018. **13**: p. 1-16.
32. Kieren, S., et al., *A biogeographic and ecological perspective to the evolution of reproductive behaviour in the family Salamandridae.* Molecular Phylogenetics and Evolution, 2018. **121**: p. 98-109.
33. Rancilhac, L., et al., *Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders.* Molecular Phylogenetics and Evolution, 2021. **155**.
34. Wielstra, B. and J.W. Arntzen, *Description of a new species of crested newt, previously subsumed in Triturus ivanbureschi (Amphibia: Caudata: Salamandridae)*. Zootaxa, 2016. **4109**(1): p. 73-80.
35. Wielstra, B., D. Salvi, and D. Canestrelli, *Genetic divergence across glacial refugia despite interglacial gene flow in a crested newt.* Evolutionary Biology, 2021. **48**(1): p. 17-26.
36. Arntzen, J.W., *Morphological and genetic diversification of pygmy and marbled newts, with the description of a new species from the wider Lisbon Peninsula (Triturus, Salamandridae)*. Contributions to Zoology, 2024. **93**(2): p. 178-200.
37. Wielstra, B., et al., *The distribution of the crested and marbled newt species (Amphibia: Salamandridae: Triturus) – an addition to the New Atlas of Amphibians and Reptiles of Europe.* Amphibia-Reptilia, 2014. **35**(3): p. 376-381.
38. Fahrbach, M. and U. Gerlach, *The genus Triturus: History, Biology, Systematics, Captive Breeding.* 2018.
39. Vucic, T., et al., *A staging table of Balkan crested newt embryonic development to serve as a baseline in evolutionary developmental studies.* Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 2024.
40. Meilink, W.R.M., et al., *Determining zygosity with multiplex Kompetitive Allele-Specific PCR (mxKASP) genotyping.* bioRxiv, 2024: p. 2024.10.25.620256.
41. De Visser, M.C., et al., *PAV-spotter: using signal cross-correlations to identify Presence/Absence Variation in target capture data* bioRxiv, 2024: p. 2024.10.25.620064.
42. Wielstra, B., et al., *Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle.* Molecular Phylogenetics and Evolution, 2019. **133**: p. 120-127.
43. Kazilas, C., et al., *Spatial genetic structure in European marbled newts revealed with target enrichment by sequence capture.* Molecular Phylogenetics and Evolution, 2024. **194**: p. 108043.
44. De Visser, M.C., et al., *NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts)*. bioRxiv, 2024: p. 2024.10.25.620290.
45. Wick, R.R., L.M. Judd, and K.E. Holt, *Performance of neural network basecalling tools for Oxford Nanopore sequencing.* Genome Biology, 2019. **20**(1): p. 129.
46. Wick, R. and J. Volkening, *Porechop: adapter trimmer for Oxford Nanopore reads.* Available online: github com/rrwick/Porechop, 2018.
47. Shafin, K., et al., *Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes.* Nature Biotechnology, 2020. **38**(9): p. 1044-1053.
48. Guan, D., et al., *Identifying and removing haplotypic duplication in primary genome assemblies.* Bioinformatics, 2020. **36**(9): p. 2896-2898.
49. Alonge, M., et al., *Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing.* Genome Biology, 2022. **23**(1): p. 258.
50. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.
51. Robinson, J.T., et al., *Integrative genomics viewer.* Nature Biotechnology, 2011. **29**(1): p. 24-6.
52. Brown, T., et al., *Sequencing and chromosome-scale assembly of the giant Pleurodeles waltl genome.* bioRxiv, 2023: p. 2022.10.19.512763.

53. Ortiz, E.M., *vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis*. 2019.
54. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*. Bioinformatics, 2014. **30**(9): p. 1312-3.
55. Letunic, I. and P. Bork, *Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool*. Nucleic Acids Research, 2024.
56. Danecek, P., et al., *Twelve years of SAMtools and BCFtools*. Gigascience, 2021. **10**(2).
57. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data*. Bioinformatics, 2011. **27**(21): p. 2987-93.
58. Danecek, P., et al., *The variant call format and VCFtools*. Bioinformatics, 2011. **27**: p. 2156-2158.
59. Smith, S.A. and B.C. O'Meara, *treePL: divergence time estimation using penalized likelihood for large phylogenies*. Bioinformatics, 2012. **28**(20): p. 2689-90.
60. Patterson, N., et al., *Ancient admixture in human history.* Genetics, 2012. **192**(3): p. 1065-93.
61. Malinsky, M., M. Matschiner, and H. Svardal, *Dsuite - Fast D-statistics and related admixture evidence from VCF files.* Molecular Ecology Resources, 2021. **21**(2): p. 584-595.
62. Pickrell, J.K. and J.K. Pritchard, *Inference of population splits and mixtures from genome-wide allele frequency data*. PLoS Genetics, 2012. **8**(11): p. e1002967.
63. Kalaentzis, K., et al., *Phylogenomics resolves the ancient and rapid radiation of banded newts (genus Ommatotriton)*. SSRN Preprints, 2024.
64. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. Gigascience, 2015. **4**: p. 7.
65. Fitak, R.R., *OptM: estimating the optimal number of migration edges on population trees using Treemix*. Biology Methods & Protocols, 2021. **6**(1): p. bpab017.
66. Berdan, E.L., et al., *Muller's ratchet and the long-term fate of chromosomal inversions*. bioRxiv, 2019: p. 606012.
67. Sessions, S.K., *Evolutionary cytogenetics in salamanders*. Chromosome Research, 2008. **16**(1): p. 183-201.
68. Kent, T.V., J. Uzunovic, and S.I. Wright, *Coevolution between transposable elements and recombination*. Philosophical Transactions of the Royal Society B, 2017. **372**(1736).
69. McCartney-Melstad, E., G.G. Mount, and H.B. Shaffer, *Exon capture optimization in amphibians with large genomes*. Molecular ecology resources, 2016. **16**: p. 1084-1094.
70. Mirarab, S., et al., *A region of suppressed recombination misleads neoavian phylogenomics*. Proceedings of the National Academy of Sciences, 2024. **121**(15): p. e2319506121.
71. Zhang, D., et al., *Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow*. Systematic Biology, 2021. **70**(5): p. 961-975.
72. Stankowski, S., et al., *The genetic basis of a recent transition to live-bearing in marine snails*. Science, 2024. **383**(6678): p. 114-119.

Lab life

Chapter 7 - General discussion

Manon de Visser[1,2]

1. Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands
2. Understanding Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands

This Chapter concludes my genomic studies of chromosome 1 syndrome in *Triturus*. The studies conducted at the Wielstra lab are part of the first research project that uses genomic information of *Triturus* newts with the aim of unraveling the evolution of balanced lethal systems in nature. Furthermore, the insights gained, as well as the methodologies and tools developed in the process, have a broader utility and thus a wider scientific impact. Therefore, I divide this Chapter into two parts. In the first part, I briefly discuss the new insights into the evolution of the balanced lethal system in *Triturus* provided by this dissertation, and I suggest avenues for further research. In the second part, I discuss the broader ways in which this dissertation contributes to science (and science education).

Part I – Balanced Lethal Systems and Further Research

For years on end, my colleagues and I have hypothesized that a balanced lethal system such as chromosome 1 syndrome in *Triturus* consists of two supergene versions with unique, recessive lethal alleles [1, 2] that;

1) Originated via a slow 'degradation process', driven by heterozygote advantage, balancing selection, and mutation accumulation due to a lack of recombination (see **Chapters 1, 2 and 6**), and/or;
2) Became established as one of the two versions – both uniquely involved in different species – entered an ancestral genome via introgressive hybridization, which facilitated fixation in the population (see **Chapters 1 and 6**).

Both of these theories are broadly applicable to diploid organisms, with the first (i.e. the slow degradation process for two supergene variants that got caught in a balanced polymorphism) gaining attention in recent population modeling studies [3, 4].

Much to our surprise, the evidence uncovered in my research does not support this gradual degradation model, but instead pointed toward a rapid – practically instantaneous – origin of the balanced lethal system in *Triturus* [see Chapter 6 and; 5]. This idea had already been suggested by researchers over 35 years ago*: "We postulate that the chromosome 1 arrest syndrome is the result of a cytogenetic accident involving an unequal genic exchange between the homologues of chromosome 1 in the common ancestor … Such an exchange could potentially result in the formation of a balanced lethal chromosomal heteromorphism in one step if the exchange: (1) was large enough to abolish homologous pairing and crossing over in the exchanged regions, (2) took place in the germ line, and (3) the exchanged material included DNA sequences that were important in early embryonic development*" [quoted from; 6]. However, these researchers never had the empirical evidence to support this claim – something that I have now co-produced.

Furthermore, with the second theory (i.e. introgressive hybridization brought both chromosomes, 1A and 1B, together in one ancestral genome) I expected to find that only the 1A-linked genes, or only the 1B-linked genes, would show evidence of a distinct evolutionary history. Instead, we not only found this for both sets of genes, but also for chromosome 1 as a whole (i.e., also including the colinear recombining region), complicating the evolutionary narrative even further (see **Chapter 6**). While we have demonstrated that introgressive hybridization likely played a role in the aberrant evolution of chromosome 1 overall, the extent of its involvement in the formation of the balanced lethal system remains unclear.

In short, this dissertation thus offers some of the first, big puzzle pieces necessary for understanding the evolutionary origin of the balanced lethal system that is chromosome 1 syndrome. Also, **Chapter 6** of this dissertation introduces the term 'bloopergene' into the scientific literature – highlighting that, despite their name, supergenes do not always result in 'superlative' outcomes for a species. However, this dissertation also demonstrates that scientific inquiry is an ongoing process, and that in relation to the origin of balanced lethal systems – both in general, as well as for the case in *Triturus* specifically – many unanswered questions remain. These can be solved in the future by applying more (elaborate) techniques.

Follow-up research: Whole Genome Studies

Firstly, while the 'NewtCap' target capture methodology (see **Chapter 4**) provides a large amount of highly insightful data, it only captures a part of the coding regions of the DNA of *Triturus*. As a result, we so far still lack a comprehensive list of genes that show (presence/absence) variation associated with the balanced lethal system for each of the affected species. Secondly, while the structural variation and gene order we hypothesize for chromosomes 1A and 1B are likely accurate given the high synteny among species used in the comparative genomic analysis [see the ancestral constitution posed in **Chapter 6** and see; 5], it is worth noting that this remains an assumption at this point. This means that our 'next-generation sequencing' techniques have proven useful in providing us with certain 'pieces of the puzzle', however it also means that the next step would be to obtain *all* the pieces, i.e. to sequence the entire genome. This can be done by performing so-called 'third-generation sequencing' techniques [7].

Future studies utilizing third-generation sequencing techniques such as whole genome sequencing of long reads – something that is currently in progress at the Wielstra lab – will help clarify most issues. I anticipate that these studies will confirm the findings presented in this dissertation. Ideally, third-generation-based whole genome sequencing will soon yield a high quality, chromosome-level reference assembly for *Triturus*, in order to further support re-sequencing projects (which could include both third-generation-

based whole genome sequencing approaches, as well as next-generation-based whole genome sequencing approaches). Once this is achieved, whole genome sequencing studies could provide a much more accurate and complete picture of *Triturus* genomes for each species, making a more detailed ancestral reconstruction of chromosome 1 possible, which in turn would help explain the evolutionary origin of it. The genotyping methods introduced in **Chapter 3** will likely be a valuable resource to support the upstream laboratory work required for executing such future whole genome sequencing projects efficiently and at the lowest costs possible.

Follow-up research: Transcriptomics and Embryology

While this dissertation does not extensively focus on the specific causes of developmental arrest in 50% of the *Triturus* embryos, it is worth mentioning that our work opens new research avenues to delve deeper into the biology of the disease itself. As described in **Chapter 1**, embryonic arrest occurs during the phylotypic stage – a period that is highly conserved in terms of both morphology and underlying, regulatory pathways. This suggests that the candidate lethal genes involved in the disrupted embryonic development could belong to the so-called 'evo-devo gene toolkits' [8] – genes that are required for basic, embryonic modules or their regulation, such as the Hox genes that determine the body plan organization. Alternatively, the candidate lethal genes could also simply be essential 'house-keeping' genes, for instance [9].

Although my co-workers did try to analyze the main functionality of the 1A- and 1B-linked targets through a gene ontology analysis [5], no pattern so far stands out. Thus, it remains to be determined precisely what role the 1A- and 1B-linked genes play in causing chromosome 1 syndrome, as well as how they interact with each other and with other genes, in terms of cellular components, molecular functions, and biological processes.

Currently, researchers of the Wielstra lab have started conducting transcriptomic studies to examine the gene expression patterns at different embryonic stages that lead up to the deadly late tail-bud stage per each embryonic class (healthy 1A1B/1B1A, diseased 1A1A, or diseased 1B1B). These studies do not only aim to compare the expression patterns between 1A- and 1B-linked genes (as essentially the genes responsible for lethality should differ between the two, a prerequisite of a balanced lethal system as also shown by my results in this dissertation), but also across all *Triturus* species. Both studies focusing on RNA, as well as embryological studies focused on morphology, will help pinpoint the specific genes and molecular pathways involved in the *Triturus* balanced lethal system, adding an 'evo-devo' perspective on the subject at hand.

With such future studies, it is important to always keep in mind that ever since the balanced lethal system became fixed in the ancestor of *Triturus*, natural selection would not have been able to eliminate new harmful (including lethal) alleles that cause earlier arrest (as the fitness cannot drop below zero – which is why I referred to the balanced lethal system as the ultimate evolutionary failure in **Chapter 1**). Given this, there likely are differences between the species – although we now have reason to believe that the effect of this 'ongoing decay' has, surprisingly, been relatively minor (see **Chapter 6**). As with whole genome sequencing studies in the future, I foresee that for transcriptomic and embryological testing particularly **Chapter 3** (on classifying embryo's in the laboratory) will prove helpful.

Part II – The Broader Impact of this Dissertation

Apart from delivering new insights into the evolutionary origin of the *Triturus* balanced lethal – an inherently compelling topic – this dissertation holds a broader relevance in both science and education. Namely, I believe that understanding the evolution of supergenes and bloopergenes in nature is not only fundamentally important, but can also indirectly have more practical implications. Here, I am not referring to stopping, or reversing, the balanced lethal system in *Triturus* embryos – after all, their ancestors thrived and speciated despite their evolutionary constraints. Rather, the insights gained from studying this system can be useful for understanding and addressing various other natural phenomena that share some of the key features of balanced lethal systems, particularly in fields where managing genetic diversity, or mitigating the effects of harmful alleles, is more relevant. Thus, I highlight some noteworthy parallels below. Finally, I will expand the scope even further by discussing how the methods and tools developed for this dissertation have the potential to advance future research projects, especially studies related to genomics and biodiversity (with a focus on salamanders).

From sickle-cell anemia in humans to hybrid vigor in maize

A balanced lethal system represents the most extreme form of heterozygote advantage possible, as all homozygotes inevitably die before reaching a reproductive age. However, more cases exist where heterozygotes that carry certain dominant alleles alongside disease-causing, recessive alleles, have a higher fitness compared to both types of homozygotes that carrying two similar copies of those harmful alleles. Thus, understanding the evolutionary mechanisms that allow harmful alleles to evolve and persist in populations could deepen our understanding of balancing selection in other contexts as well.

Sickle-cell anemia poses an example. On the one hand, humans that are homozygous for the 'mutated' hemoglobin allele (HbS/HbS) suffer from sickle-cell anemia: a condition in which red blood cells deform, leading to serious health problems and sometimes death [10-12]. On the other hand, if a person is homozygous for the 'normal' hemoglobin allele (HbA/HbA) they will not develop this heritable disease – however they are fully susceptible to severe malaria, caused by *Plasmodium falciparum* [13]. Intruigingly, individuals that are heterozygous for the sickle-cell-associated trait (HbA/HbS), and that thus carry one normal and one mutated hemoglobin allele, are protected against the malaria parasite *without* developing the severe symptoms of sickle-cell anemia [14-16]. Thus, balancing selection maintains the harmful hemoglobin allele in human populations that occur in areas with a high prevalence of malaria due to heterozygote advantage, as both types of homozygotes are less fit [15, 17, 18]. In this case, the outcome is of course heavily influenced by environmental factors – which is not applicable (anymore) to balanced lethal systems – but the driving, evolutionary mechanisms are similar nonetheless. Essentially, this example of balancing selection in humans demonstrates the same thing the *Triturus* balanced lethal system teaches: nature does not always offer an 'easy way out', as natural selection is never able to anticipate future conditions.

Another relevant context is that of plant cultivation practices in agricultural science. For instance, hybrid vigor (another word for heterozygote advantage) can be used to improve crops, such as corn [19]. When two different, homozygous (inbred) lines are crossed, this can result in heterozygous (hybrid) offspring that outperform the parent lines in terms of fitness. This can lead to enhanced yield and disease resistance, for example [20]. In fact, the word 'Hi-Bred' that later became the more general word 'hybrid' was first used by a company in 1926 in the context of crop improvement by applying hybrid vigor through crossing different breeds [21]. For corn specifically, hybrid breeds can show a yield increase of 30% compared to non-hybrid varieties [22] and they are more resistant to diseases like stalk rot [23] and leaf blight [24]. Also, corn hybrids may be more equipped to deal with external stressors, such as drought conditions [25] – something that is of importance in regions affected by climate change or inconsistent rainfall patterns. Thus, by understanding how balanced lethal systems evolve naturally, insights can be obtained that can in turn be used to study and improve hybrid vigor artificially. This will lead to higher food security and increased sustainability – which are of economic and environmental importance.

A helicopter view

My research primarily addresses a fundamental question within evolutionary biology. But no research is ever truly fundamental, as methods and tools developed for fundamental research often find applications in unrelated fields. This dissertation is an example of that:

combined, **all Chapters** form a coherent story describing all the chronological steps that I took over the course of five years to study the *Triturus'* balanced lethal system with available molecular techniques. But most of the work conducted holds potential for applications beyond this original research scope as well. For the final part of this Chapter, I would thus like to zoom out to overlook certain elements that have been briefly touched upon in other parts of this dissertation – and while doing so I will elaborate on the three main topics that personally drive me as an applied scientist.

Animal Welfare

"The least I can do is speak out for those who cannot speak for themselves" – Dr. Jane Goodall

Surprisingly, this dissertation could offer an innovative solution to some pressing animal welfare concerns in the intensive farming industry. This topic relates specifically to **Chapter 3**, where we discuss the potential of our mxKASP method for sexing embryos while still inside the egg (also referred to as '*in ovo* sex determination'). Worldwide, this practice is investigated in order to address ethical and economic issues, particularly in the context of male chick culling in the poultry industry. While our mxKASP approach was originally developed to differentiate between the genotypes of diseased and healthy *Triturus* embryos (see **Chapter 3**), it can certainly be adapted for broader use. Although implementing mxKASP to genotype chicken embryos at embryonic stages in which pain perception is physiologically implausible at a large scale would involve numerous steps, it is crucial to realize the importance of cultivating novel ideas that can sometimes emerge from unrelated research efforts.

Annually, it is estimated that seven *billion* one-day-old male chicks are slaughtered in the poultry industry worldwide [26], a staggering number that has likely increased since it was reported over six years ago due to the continuous growth of the industry (see Production of eggs worldwide 2022 | Statista). Common methods for slaughtering – or 'culling' – male chicks include gassing and maceration. With gassing, high concentrations of carbon dioxide are used to kill the chicks, however this can cause hypercapnia (excessive carbon dioxide in the bloodstream) and hypoxia (a deprivation of oxygen) before individuals lose consciousness, meaning that this approach – although generally considered humane – can still inflict pain and suffering [27]. Furthermore, maceration is a way to mechanically decimate the chicks, mostly using a 'chick shredder': a high-speed machine with rotating blades [28]. Although generally perceived as cruel by the public, this method – if performed properly – is more fast and painless as compared to gassing [27, 29, 30]. However, the execution of chick shredding can be far from proper, and

hazards with serious, negative welfare consequences are identified in practice, like; blades or rollers that are rotating too slowly, machines that are overloaded (i.e., too many chicks are inserted at once), or rollers that are spaced too widely [31]. These hazards can lead to large numbers of chicks not being killed instantaneously, and thus experiencing prolonged pain, fear, and distress.

Several techniques are under investigation for *in ovo* sexing of chicken eggs, and two major categories of approaches are the spectrometry-based techniques and the molecular techniques. On the one hand, spectrometry-based techniques are relatively fast and do not require DNA extraction, but involve complex data processing steps and can lack accuracy in some cases [32, 33]. On the other hand, molecular approaches, such as genotyping, are highly accurate but require more time and detailed laboratory procedures [32, 34]. The mxKASP set-up, as discussed in **Chapter 3**, would include a DNA extraction step, but eliminates the need for post-PCR analyses (such as checking results on gel, as is the case in standard PCR). Furthermore, KASP is extremely well-suited for high-throughput genotyping, which is essential for a rapid, large-scale usage in farming.

Wildlife Conservation

"In the end, it's not just about the animals, it's about us and our survival as well." – Steve Irwin

The biodiversity of the world is currently declining rapidly in what is already referred to as the 'sixth mass extinction' [35]. For the sake of simplicity and continuity, I will here focus specifically on the decline of amphibian populations. As discussed in **Chapter 1**, amphibians represent the most threatened group of vertebrates and they are, unfortunately, easily affected by anthropogenic impacts on their environments [36-39]. It is estimated that over 70% of the world's amphibians are currently experiencing population declines, primarily due to human colonization and globalization [36, 40-42]. The six main drivers behind their mass disappearance include; emerging diseases, habitat loss, pollution, climate change and weather extremes, the illegal pet trade, and threats caused by invasive species [37-39, 43-45].

Chapter 4 of this dissertation is primarily relevant to amphibian wildlife conservation, as it highlights a novel and detailed approach that makes it possible to investigate, monitor, and conserve a specific group of amphibians worldwide. Initially, we optimized the target capture protocol (now termed 'NewtCap') for studying the balanced lethal system in *Triturus* newts specifically. However the idea to simultaneously test how effective the approach is for studying other research questions in other salamander species appeared a worthwhile pursuit – not primarily for my personal research output,

but more so in terms of future impact. As demonstrated in **Chapter 4**, NewtCap data have proven useful in a broad range of research fields, including conservation genetics.

In discussing the future of studying and protecting amphibians, I want to emphasize their critical role as bioindicators in ecosystems. Firstly, they form an important link in food chains worldwide [46, 47]. Secondly, they are particularly vulnerable to for instance diseases, pollution, and drought due to their permeable skin and limited dispersal capabilities [48-51]. Simply put: the more biodiverse and natural the environment is, the more likely it is to support healthy and diverse amphibian populations – hence the term 'bioindicator'. As we continue to lose amphibians at an unprecedented rate, it should serve as a warning as a healthy environment is of importance for our own survival as well. Sir David Attenborough mentioned the following on the matter, back in 2008: *"Amphibians are the lifeblood of many environments, playing key roles in the function of ecosystems, and it is both extraordinary and terrifying that in just a few decades the world could lose half of all these species."* For me personally, it has been a privilege that the *Triturus* project provided me the opportunity to contribute to a study that can be of great importance to the preservation of a wide range of salamanders across the globe.

Science Communication

"If enlightenment was the salad, entanglement is the soup" – Dr. Neri Oxman

The creation of this dissertation has only been possible thanks to the numerous collaborations with; 1) other academic research institutes, 2) specialized (non-academic) research organizations, and 3) citizen scientists and volunteers. In this section, I want to primarily focus on the importance of collaborating and communicating beyond academia (i.e., number 2 and 3). For instance, the *Triturus* embryos that I analyzed in the context of the balanced lethal system – and that thus form the basis for **Chapter 3**, as well as the empirical research in **Chapter 6** – were collected by a hobbyist breeder. Similarly, some of the samples provided to re-build the Salamandridae phylogeny of **Chapter 4** also originated from cooperative hobbyists and pet owners. Additionally, I have, for instance, co-supervised and co-authored student projects that were not part of this dissertation, but that did yield remarkable outcomes for the herpetological sciences [52-66]. Again, these projects were partially made possible through the support of motivated volunteers, as well as practitioners, who assisted with fieldwork and with the publication process. This clearly demonstrates the importance of working together with citizen scientists, hobbyists, and other types of professional experts outside of the academic 'bubble'.

Non-academic experts and hobbyists often do not only have a deep motivation to help science progress, but they also tend to bring substantial 'niche knowledge' that can

significantly enhance the scientific process [67-69]. And people outside of expert circles may be unaware of the studies or research projects conducted by scientific institutes, but they often are eager to learn and become engaged once introduced to a certain subject [70]. Moreover, I strongly believe that the general public can assist scientists in translating their research to broader audiences more effectively (as they themselves are part of that audience). For example, one article I co-authored a few years ago (which is, again, not a part of this dissertation, but which is a part of the overall balanced lethal system project of our research team) got published in *Frontiers for Young Minds*: a scientific journal that is peer-reviewed by highly capable and motivated teenagers [2]. Similarly, **Chapter 2** of this dissertation is based on a popular science article that I wrote to make the balanced lethal system topic accessible not only to the scientific community, but also to experts, volunteers, and hobbyists interested in the aberrant evolution and genetics of crested and marbled newts [71]. While writing it, I specifically consulted non-biologist friends.

Unfortunately, within academia, conducting scientific outreach has traditionally been – and often still is – perceived as a low-status task, usually delegated to or reserved for graduate students and early-career scientists [predominantly women; 72, 73, 74]. Personally, I find engaging with broader audiences not only highly rewarding, but I also consider it an ethical responsibility to educate the public. Whether scientists themselves should be responsible for public engagement remains a topic of debate – and similar debates exist regarding what are the best approaches for doing so [75-77]. Also, similar debates apply to determining how best to direct the future of science. For instance, there is much discussion around 1) how to fairly and effectively evaluate and promote academic performance [78-80], and 2) how to balance academic research and scientific integrity with the increasing interconnectedness between science, industry, and society [81, 82]. And while these issues are beyond the scope of this dissertation, I will offer the following, given my active involvement in public engagement (see Appendices):

Regardless of how the academic landscape evolves; conducting solid research will always be essential. It leads to robust, high-quality science – an intriguing salad. But establishing strong collaborations and communicating science effectively? That is what makes for a big cup of warm soup.

References

1. Wielstra, B., *Balanced lethal systems.* Current Biology, 2020. **30**: p. R742-R743.
2. Meilink, W.R.M., et al., *Balanced Lethal Systems: an evolutionary mystery.* Frontiers for Young Minds, 2021. **9**.
3. Berdan, E.L., et al., *Muller's ratchet and the long-term fate of chromosomal inversions.* bioRxiv, 2019: p. 606012.
4. Berdan, E.L., et al., *Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals.* Philosophical Transactions of the Royal Society B, 2022. **377**(1856): p. 20210199.
5. France, J., et al., *Genomic evidence suggests the balanced lethal system in Triturus newts originated in an instantaneous speciation event.* bioRxiv, 2024: p. 2024.10.29.620207.
6. Sessions, S.K., et al., *Cytology, embryology, and evolution of the developmental arrest syndrome in newts of the genus Triturus (Caudata: Salamandridae).* The Journal of Experimental Zoology, 1988. **248**: p. 321-334.
7. Giani, A.M., et al., *Long walk to genomics: History and current approaches to genome sequencing and assembly.* Computational and Structural Biotechnology Journal, 2020. **18**: p. 9-19.
8. Friedrich, M., *Evo-Devo gene toolkit update: At least seven Pax transcription factor subfamilies in the last common ancestor of bilaterian animals.* Evolution and Development, 2015. **17**: p. 255-257.
9. Joshi, C.J., et al., *What are housekeeping genes?* PLoS Computational Biology, 2022. **18**(7): p. e1010295.
10. Steinberg, M.H., *Sickle Cell Anemia, the First Molecular Disease: Overview of Molecular Etiology, Pathophysiology, and Therapeutic Approaches.* The Scientific World JOURNAL, 2008. **8**: p. 1295-1324.
11. Herrick, J.B., *Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia.* Archives of internal medicine, 1910. **6**(5): p. 517-521.
12. Pauling, L., et al., *Sickle cell anemia, a molecular disease.* Science, 1949. **110**(2865): p. 543-548.
13. Snow, R.W., et al., *The global distribution of clinical episodes of Plasmodium falciparum malaria.* Nature, 2005. **434**(7030): p. 214-217.
14. Allison, A.C., *The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1954. **48**(4): p. 312-318.
15. *A novel locus of resistance to severe malaria in a region of ancient balancing selection.* Nature, 2015. **526**(7572): p. 253-257.
16. Band, G., et al., *Malaria protection due to sickle haemoglobin depends on parasite genotype.* Nature, 2021. **602**(7895): p. 106-111.
17. Elguero, E., et al., *Malaria continues to select for sickle cell trait in Central Africa.* Proceedings of the National Academy of Sciences, 2015. **112**(22): p. 7051-7054.
18. Habibzadeh, F., *The effect on the equilibrium sickle cell allele frequency of the probable protection conferred by malaria and sickle cell gene against other infectious diseases.* Scientific Reports, 2024. **14**(1).
19. Hochholdinger, F. and J.A. Baldauf, *Heterosis in plants.* Current Biology, 2018. **28**(18): p. R1089-R1092.
20. Liu, J., et al., *Exploring the molecular basis of heterosis for plant breeding.* Journal of Integrative Plant Biology, 2020. **62**(3): p. 287-298.
21. Bruns, H.A., *Southern Corn Leaf Blight: A Story Worth Retelling.* Agronomy Journal, 2017. **109**(4): p. 1218-1224.
22. Sheoran, S., et al., *Recent Advances for Drought Stress Tolerance in Maize (Zea mays L.): Present Status and Future Prospects.* Frontiers in Plant Science, 2022. **13**: p. 872566.
23. Zhu, M., et al., *Genetic dissection of maize disease resistance and its applications in molecular breeding.* Molecular Breeding, 2021. **41**(5): p. 32.
24. Abera, W., et al., *Heterosis and combining ability of elite maize inbred lines under northern corn leaf blight disease prone environments of the mid-altitude tropics.* Euphytica, 2015. **208**(2): p. 391-400.
25. Ahmad, Z., et al., *Studies on screening of maize (Zea mays L.) hybrids under drought stress conditions.* Journal of Advance Botany and Zoology, 2015. **2**: p. 1-5.

26. Krautwald-Junghanns, M.E., et al., *Current approaches to avoid the culling of day-old male chicks in the layer industry, with special reference to spectroscopic methods.* Poultry Science, 2018. **97**(3): p. 749-757.
27. Gurung, S., et al., *Evaluation of Alternative Euthanasia Methods of Neonatal Chickens.* Animals (Basel), 2018. **8**(3).
28. Rutt, R.L. and J. Jakobsen, *The 'brother layer problem': Routine killing, biotechnology and the pursuit of 'ethical sustainability' in industrial poultry.* Environment and Planning E: Nature and Space, 2022. **6**(3): p. 1785-1803.
29. Wang, X., et al., *Evaluation of Euthanasia Methods on Behavioral and Physiological Responses of Newly Hatched Male Layer Chicks.* Animals (Basel), 2021. **11**(6).
30. van Niekerk, T. and J. Workamp, *Scenarios for addressing the dilemma of 'the culling of day-old male chicks of layerbreeds'*. 2022, Wageningen Livestock Research.
31. AHAW, E.P.o.A.H.a.W., et al., *Killing for purposes other than slaughter: poultry.* EFSA Journal, 2019. **17**(11): p. e05850.
32. Jia, N., et al., *A Review of Key Techniques for in Ovo Sexing of Chicken Eggs.* Agriculture, 2023. **13**(3).
33. Galli, R., et al., *In ovo sexing of chicken eggs by fluorescence spectroscopy.* Analytical and Bioanalytical Chemistry, 2017. **409**(5): p. 1185-1194.
34. Corion, M., et al., *Trends in in ovo sexing technologies: insights and interpretation from papers and patents.* Journal of Animal Science and Biotechnology, 2023. **14**(1): p. 102.
35. Ceballos, G., et al., *Accelerated modern human–induced species losses: Entering the sixth mass extinction.* Science advances, 2015. **1**(5): p. e1400253.
36. Stuart, S.N., et al., *Status and trends of amphibian declines and extinctions worldwide.* Science, 2004. **306**: p. 1783-1787.
37. Araujo, A., et al., *Micro(nano)plastics as an emerging risk factor to the health of amphibian: A scientometric and systematic review.* Chemosphere, 2021. **283**: p. 131090.
38. Hof, C., et al., *Additive threats from pathogens, climate and land-use change for global amphibian diversity.* Nature, 2011. **480**(7378): p. 516-9.
39. Ford, J., et al., *Adrift on a Sea of Troubles: Can Amphibians Survive in a Human-Dominated World?1*. Herpetologica, 2020. **76**(2).
40. Hayes, T.B., et al., *The cause of global amphibian declines: a developmental endocrinologist's perspective.* Journal of Experimental Biology, 2010. **213**: p. 921-933.
41. Sparreboom, M., *Salamanders of the Old World: the salamanders of Europe, Asia and northern Africa*. 2014: p. 431p.
42. Pabijan, M., et al., *Evolutionary principles guiding amphibian conservation.* Evolutionary Applications, 2020. **00**: p. 1-22.
43. Hughes, A.C., B.M. Marshall, and C.T. Strine, *Gaps in global wildlife trade monitoring leave amphibians vulnerable.* eLife, 2021. **10**.
44. Stringham, O.C., J.L. Lockwood, and C. Bellard, *Pet problems: Biological and economic factors that influence the release of alien reptiles and amphibians by pet owners.* Journal of Applied Ecology, 2018. **55**(6): p. 2632-2640.
45. Lucas, P.M., et al., *Using comparative extinction risk analysis to prioritize the IUCN Red List reassessments of amphibians.* Conservation Biology, 2024: p. e14316.
46. West, J., *Importance of amphibians: A synthesis of their environmental functions, benefits to humans, and need for conservation*. 2018.
47. Gobel, N., et al., *Invariant and vulnerable food web components after bullfrog invasion.* Biological Invasions, 2022. **25**(3): p. 901-916.
48. Valencia-Aguilar, A., A.M. Cortés-Gómez, and C.A. Ruiz-Agudelo, *Ecosystem services provided by amphibians and reptiles in Neotropical ecosystems.* International Journal of Biodiversity Science, Ecosystem Services & Management, 2013. **9**(3): p. 257-272.
49. Hocking, D.J. and K.J. Babbitt, *Amphibian contributions to ecosystem services.* Herpetological conservation and biology, 2014.
50. Van der Goot, A., M. De Visser, and A.F. Hiemastra, *Smooth newts Lissotriton vulgaris observed hibernating in a waterfowl nest.* Herpetological Bulletin, 2022. **162**: p. 41-42.
51. Saber, S., W. Tito, and R. Said, *Amphibians as Bioindicators of the Health of Some Wetlands in Ethiopia*. The Egyptian Journal of Hospital Medicine, 2017. **66**: p. 66-73.

52. Kalaentzis, K., et al., *Hybrid zone analysis confirms cryptic species of banded newt and does not support competitive displacement since secondary contact.* Ecology and Evolution, 2023. **13**(9): p. e10442.
53. Kazilas, C., et al., *Spatial genetic structure in European marbled newts revealed with target enrichment by sequence capture.* Molecular Phylogenetics and Evolution, 2024. **194**: p. 108043.
54. Struijk, R.P., et al., *Knoflookpad in Callantsoog: Verspreiding, voortplanting en geografische herkomst.* RAVON, 2024. **26**(1): p. 2-5.
55. Kuijt, M., et al., *The introduction of three cryptic tree frog species in the Dutch coastal dunes challenges conservation paradigms.* Amphibia-Reptilia, 2022. **44**(1): p. 1-10.
56. Koster, S., et al., *The conservation paradox of an introduced population of a threatened species: spadefoot toads in the coastal dunes of the Netherlands.* Amphibia-Reptilia, 2022. **44**(1): p. 11-18.
57. de Brouwer, J., et al., *An isolated crested newt population in Dutch coastal dunes: distribution relict or introduction?* Amphibia-Reptilia, 2022. **44**(1): p. 19-26.
58. Vliegenthart, C., et al., *The limits of mtDNA analysis for determining the provenance of invasive species: a midwife toad example.* Amphibia-Reptilia, 2022. **44**(1): p. 27-33.
59. De Visser, M., et al., *Exotische amfibieën in de duinen ontmaskerd met mtDNA barcoding.* Holland's Duinen, 2023. **82**(1): p. 25-29.
60. Fahrbach, M., M. de Visser, and B. Wielstra, *The hybrid zone between the Italian and Northern Crested Newts (Triturus carnifex and T. cristatus) reaches Germany.* Salamandra, 2021. **57**(3).
61. Struijk, R.P., et al., *'Rise and fall'van een populatie westelijke smaragdhagedissen in Scheveningen.* RAVON, 2022. **24**(2): p. 24-27.
62. Struijk, R.P., M. de Visser, and B. Wielstra, *Exotische smaragdhagedissen in de duinen bij Scheveningen.* Kijk op Exoten, 2021. **9**(3): p. 12-13.
63. Wielstra, B., et al., *Herkomst van exotische Alpenwatersalamanders blootgelegd.* Schubben & Slijm, 2023. **54**: p. 8-9.
64. Wielstra, B., et al., *MtDNA barcoding van exotische amfibieën in de duinen.* RAVON, 2023. **25**(2): p. 26-29.
65. Mars, J., et al., *Phylogenomics yields new systematic and taxonomical insights for Lissotriton newts, a lineage with a strong legacy of introgressive hybridization.* SSRN Preprints, 2024.
66. Robbemont, J., et al., *An extended mtDNA phylogeography for the alpine newt illuminates the provenance of introduced populations.* Amphibia-Reptilia, 2023. **44**(3): p. 347-361.
67. Jones, M.G., et al., *Citizen scientists and non-citizen scientist hobbyists: Motivation, benefits, and influences.* International Journal of Science Education, Part B, 2018. **8**(4): p. 287-306.
68. Kobori, H., et al., *Citizen science: a new approach to advance ecology, education, and conservation.* Ecological Research, 2015. **31**(1): p. 1-19.
69. Bonney, R., et al., *Can citizen science enhance public understanding of science?* Public Understanding of Science, 2016. **25**(1): p. 2-16.
70. Jensen, E. and N. Buckley, *Why people attend science festivals: Interests, motivations and self-reported benefits of public engagement with research.* Public Underst Sci, 2014. **23**(5): p. 557-73.
71. De Visser, M., et al., *Een evolutionair raadsel: het dodelijke chromosoom 1 syndroom in Triturus-salamanders.* RAVON, 2021. **23**(1): p. 9-12.
72. Woitowich, N.C., et al., *Assessing Motivations and Barriers to Science Outreach within Academia: A Mixed-Methods Survey.* bioRxiv, 2021: p. 2021.10. 28.466319.
73. Johnson, D.R., E.H. Ecklund, and A.E. Lincoln, *Narratives of Science Outreach in Elite Contexts of Academic Science.* Science Communication, 2013. **36**(1): p. 81-105.
74. Bankston, A. and G.S. McDowell, *Changing the Culture of Science Communication Training for Junior Scientists.* Journal of Microbiology & Biology Education, 2018. **19**(1).
75. Peacock, K.A., *The Duty to Communicate.* BioScience, 2020. **70**(8): p. 719-721.
76. Gardner, C.J., et al., *From Publications to Public Actions: The Role of Universities in Facilitating Academic Advocacy and Activism in the Climate and Ecological Emergency.* Frontiers in Sustainability, 2021. **2**.
77. Llorente, C., et al., *Scientists' opinions and attitudes towards citizens' understanding of science and their role in public engagement activities.* PLoS ONE, 2019. **14**(11): p. e0224262.
78. Edwards, M.A. and S. Roy, *Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition.* Environmental Engineering Science, 2017. **34**(1): p. 51-61.

79. Kassab, O., *Does public outreach impede research performance? Exploring the 'researcher's dilemma' in a sustainability research center.* Science and Public Policy, 2019. **46**(5): p. 710-720.
80. Herbaut, E. and K. Geven, *What works to reduce inequalities in higher education? A systematic review of the (quasi-)experimental literature on outreach and financial aid.* Research in Social Stratification and Mobility, 2020. **65**.
81. Mecca, J.T., et al., *Researcher Perspectives on Conflicts of Interest: A Qualitative Analysis of Views from Academia.* Science & Engineering Ethics, 2015. **21**(4): p. 843-55.
82. Johns, M.M., M. Barnes, and P.S. Florencio, *Restoring balance to industry-academia relationships in an era of institutional financial conflicts of interest: promoting research while maintaining trust.* Journal of the American Medical Association, 2003. **289**(6): p. 741-746.

Swabbing of *Triturus cristatus*

Appendix I: Nederlandse samenvatting

Algemene inleiding

In dit proefschrift wordt het evolutionaire raadsel genaamd een 'gebalanceerd letaal systeem' onderzocht. Dit fenomeen is het best beschreven in de literatuur van kam- en marmersalamanders (oftewel, soorten van het geslacht '*Triturus*'). Het syndroom, dat ook wel 'chromosoom 1 syndroom' genoemd wordt, is door wetenschappers in verband gebracht met het bestaan van twee zeer verschillende chromosoomversies binnen chromosomenpaar 1 (het langste chromosoom van de twaalf chromosomenparen die de salamanders bezitten). Er wordt vermoed dat beide versies van chromosoom 1 – ook wel 1A en 1B genoemd – unieke, essentiële genen missen, waardoor de helft van alle bevruchte eieren al sterven nog voordat ze kunnen uitkomen, omdat nakomelingen 50% kans hebben twee keer dezelfde versie te erven (en daarmee dus automatisch de essentiële genen op de alternatieve versie missen). Zo'n gebalanceerd letaal systeem lijkt in strijd met de basisprincipes van de evolutie: omdat een syndroom dat leidt tot een structureel verlies van 50% van de reproductieve output namelijk evolutionair gezien veel kost, is de verwachting dat de werking van natuurlijke selectie het ontstaan – en het voortbestaan – daarvan nooit zou toelaten. Echter, naar alle waarschijnlijkheid houdt het natuurlijke fenomeen al minstens 24 miljoen jaar stand. Dit is de aanname, omdat een gebalanceerd letaal systeem voorkomt in alle soorten die behoren tot de *Triturus* groep: een grote aanwijzing dat het ontstaan moet zijn in de gemeenschappelijke voorouder van deze groep.

Om de evolutie van dit systeem te begrijpen, levert dit proefschrift innovatieve methodes en inzichten. De studies in dit proefschrift zijn het resultaat van twee, overkoepelende doelen; 1) het streven naar het ontdekken – en het genetisch gezien 'in kaart brengen' – van de mutaties op de 1A en de 1B versie van chromosoom 1 (oftewel, het aantonen van de aan- en afwezigheid van bepaalde, essentiële genen die betrokken zijn bij het syndroom), en 2) het gebruiken van die informatie om door middel van empirisch, moleculair onderzoek tot nieuwe inzichten te komen over de evolutionaire oorsprong van het gebalanceerde letale systeem binnen *Triturus*. Deze studies – die in het Engels zijn gerapporteerd in dit proefschrift (als 'Hoofdstuk 2' t/m 'Hoofdstuk 6'), die gepubliceerd zijn op een *preprint* server, en die op het moment van schrijven van dit proefschrift ter beoordeling zijn overhandigd aan wetenschappers via verscheidene, erkende, wetenschappelijke tijdschriften – worden hieronder verder samengevat.

Hoofdstuk 2: Een evolutionair raadsel – het dodelijke chromosoom 1 syndroom

Dit hoofdstuk introduceert het concept van het gebalanceerde letale systeem en legt uit hoe *Triturus* salamanders – ondanks het bizarre chromosoom 1 syndroom – tóch kunnen voortbestaan. Normaal gesproken worden dodelijke mutaties (zoals het 'missen' van cruciale genen) namelijk ofwel sterk onderdrukt, ofwel geheel uit populaties verwijderd. Dit komt door de kracht van natuurlijke selectie: DNA dat goed werkt en tot veel nakomelingen leidt, wordt veelal doorgegeven – terwijl DNA dat tot kwaaltjes en dus minder nakomelingen leidt, over de generaties heen vanzelf verloren gaat. Echter, omdat in een gebalanceerd, letaal systeem alleen de heterozygote individuen (oftwel, de individuen met zowel een chromosoom 1A als een chromosoom 1B versie) overleven, lijkt natuurlijke selectie hier niet toe in staat. Dit komt omdat de twee verschillende chromosoomversies elkaar als het ware 'opheffen': versie 1A bevat bepaalde genen die essentieel zijn voor overleving, terwijl versie 1B deze genen mist – en andersom. Dit betekent allereerst dat alle homozygote embryo's sterven, wat leidt tot een dramatisch verlies in reproductieve output, maar het betekent ook dat alle mutaties doorgegeven zullen blijven worden (want zowel 1A als 1B is nodig voor de overleving). Deze paradox roept de vragen op: wat zijn precies deze genen? En hoe komt zo een nadelig systeem in de natuur tot stand? Dit hoofdstuk werkt dan ook een theorie uit die uitlegt hoe chromosoom 1 syndroom op een geleidelijke manier zou kunnen ontstaan; namelijk door een graduele opeenstapeling van steeds meer mutaties in de homozygoten, wanneer heterozygoten al van een bepaald evolutionair voordeel genieten – tot het uiterlijke en onomkeerbare punt van dubbele, homozygote letaliteit (en gebalanceerde letaliteit) bereikt is. Het systeem is dan, als het ware, over de tijd heen 'vastgeroest'.

Hoofdstuk 3: mxKASP genotypering als methode om embryo's te classificeren

Om de genetische verschillen tussen de levensvatbare en de niet-levensvatbare *Triturus* embryo's te onderzoeken, is een nieuwe genotyperingsmethode ontwikkeld: multiplex Kompetitive Allele-Specific PCR (ook wel 'mxKASP'). Deze techniek maakt het mogelijk om op een snelle en grootschalige wijze vast te stellen of een *Triturus* embryo heterozygoot is en dus levensvatbaar is (d.w.z., een '1A1B' of een '1B1A' genotype heeft), of dat deze homozygoot is en dus niet-levensbatbaar is (d.w.z. een '1A1A' of een '1B1B' genotype heeft). Dit wordt gedaan aan de hand van het zichtbaar maken van zogenaamde 'merkers'. Dit zijn stukjes DNA die geselecteerd kunnen worden omdat deze iets specifieks aantonen (in dit geval is één chromosoom 1A-specifieke merker, en één chromosoom 1B-specifieke merker gebruikt). Het toepassen van mxKASP is dus een cruciale stap in het onderzoek naar het gebalanceerde letale systeem in *Triturus*. De mxKASP methode wordt in dit hoofdstuk gevalideerd aan de hand van een soortgelijke test, namelijk een 'standaard multiplex PCR'. In de praktijk, echter, is een PCR test een stuk tijdrovender en ook minder geschikt voor het genotyperen van grote aantallen

monsters tegelijk. De mxKASP methode is, naast in dit onderzoek naar chromosoom 1 syndroom, ook te gebruiken in andere genetische studies waar stukken genoom aan- of afwezig zijn. Dat kan door het gebruik van andere merkers. Zoals een studie in het kader van geslachtschromosomen en geslachtsverhoudingen binnen populaties (door in een XY-systeem, bijvoorbeeld, een X- en een Y-merker te onderzoeken).

Hoofdstuk 4: NewtCap – een efficiënte 'DNA vangst' methode voor salamanders
Omdat salamanders over extreem grote genomen (oftewel, extreem uitgebreide DNA codes) beschikken, is het analyseren van het hele genomen van meerdere individuen geen haalbare optie. Dat kan naar verwachting wel in de toekomst, maar kost vandaag de dag nog veel te veel geld, tijd, en rekenkracht van computers. Daarom is een alternatieve methode ontwikkeld waarmee bepaalde stukken uit het genoom – bijvoorbeeld genen met essentiële functies – in het specifiek worden geanalyseerd. In het Engels heet de overkoepelende methode 'target capture', omdat het gericht op zoek gaat naar stukjes ('targets') en deze 'vangt' ('capture'). Omdat deze methode zich specifiek richt op watersalamander ('newt') DNA, heeft de methode de naam 'NewtCap' gekregen. In totaal kan NewtCap informatie opleveren over zo'n zevenduizend genen, en is de methode aantoonbaar toepasbaar op alle (d.w.z., meer dan honderd) watersalamandersoorten, inclusief die van het geslacht *Triturus*. De methode is dan ook cruciaal om de eerste genetische studies naar het gebalanceerde letale systeem mogelijk te maken, maar is tegelijkertijd eigenlijk nog veelbelovender. De data die met NewtCap verkregen kunnen worden, blijken namelijk waardevol voor diverse typen onderzoeken binnen de biologie, bijvoorbeeld fylogenomische en fylogeografische studies, en vraagstukken omtrent populatiebeheer, natuurbehoud, en soortenhybridisatie. Dit is van belang, aangezien veel soorten – juist door die problematische, grote 'genomen' – nog niet altijd goed onderzocht zijn. En aangezien salamanders – net als andere amfibieën – wereldwijd steeds meer bedreigd worden met uitsterven.

Hoofdstuk 5: PAV-spotter – het identificeren van structurele genetische variatie
Over het algemeen staat 'genetische variatie' bekend als de verschillen in DNA 'letters' (nucleotiden) op bepaalde posities in de genetische code van verschillende individuen of soorten. Maar genetische variatie bestaat óók op een grotere schaal. Er wordt in dat geval van 'structurele variatie' gesproken. Een bepaalde vorm van structurele variatie is de aan- of afwezigheid van stukken DNA code (in het Engels heet dit 'Presence/Absence Variation', afgekort: 'PAV'). Dit kan gaan over bijvoorbeeld hele genen, wat ons bij het overkoepelende onderzoek naar gebalanceerde, letale systemen brengt. Om namelijk op zoek te gaan naar genen die afwezig zijn van chromosoom 1A, maar aanwezig zijn op chromosoom 1B – en andersom – is een nieuwe bioinformatica tool ontwikkeld: 'PAV-

spotter'. Deze tool gebruikt signaal-analyse om de aan- en afwezigheidsvariatie van genen in DNA data (gegenereerd met NewtCap, Hoofdstuk 4) van *Triturus* embryos (gegenotypeerd met mxKASP, Hoofdstuk 3) te detecteren. PAV-spotter blijkt succesvol in het identificeren van tientallen genen die compleet ontbreken op ófwel chromosoom 1A, ófwel chromosoom 1B. Dit bevestigt dat het gebalanceerde letale systeem wordt veroorzaakt door complementaire 'gaten' op beide chromosoomversies – en hoe dit zit wordt verder onderzocht in Hoofdstuk 6. De signaal-analyse aanpak stemt overigens uit een ander vakgebied, namelijk de elektrotechniek/akoestiek. Dat maakt dit hoofdstuk zeer multidisciplinair.

Hoofdstuk 6: De evolutionaire oorsprong van het gebalanceerde letale systeem
Dit empirische hoofdstuk biedt een aantal nieuwe inzichten in hoe het gebalanceerde letale systeem binnen *Triturus* genetisch in elkaar zit en is ontstaan. Om tot deze inzichten te komen, werden de methodes ontwikkeld in de voorgaande hoofdstukken strategisch samengebracht en ingezet. De resultaten uit dit hoofdstuk tonen aan dat de unieke 'gaten' op chromosoom 1A en 1B bij alle *Triturus* soorten vrijwel geheel consistent zijn, wat suggereert dat het systeem zich razendsnel heeft gevormd in een vroege gemeenschappelijke voorouder – en dat het daarna niet veel meer veranderd is. Dit ontkracht de theorie die uiteengezet werd in Hoofdstuk 2 dan ook direct. Daarnaast wijzen de uitkomsten van aanvullende analyses (zoals fylogenetische testen, evenals vergelijkingen met het DNA van andere watersalamandersoorten die nadrukkelijk niet lijden aan een gebalanceerd, letaal systeem) uit dat chromosoom 1 er een heel andere evolutionaire geschiedenis op nahoudt dan alle andere chromosomen uit het genoom *Triturus*-genoom. Chromosoom 1 toont namelijk een nauwe verwantschap aan met onder andere alpenwatersalamanders (van het geslacht '*Ichthyosaura'*) – in plaats van met kleine watersalamanders (van het geslacht '*Lissotriton'*), waar de rest van het genoom het nauwst verwant aan is. Dat dit zo is, komt waarschijnlijk doordat in een ver verleden verschillende soorten watersalamanders onderling hybriden vormden, en dat in dat proces een heel chromosoom werd uitgewisseld (een extreme vorm van 'introgressie'). Het is nog onduidelijk of, en hoe, dit het ontstaan van chromosoom 1 syndroom precies heeft beïnvloedt, maar het is een opvallende bevinding die op zijn minst onderstreept dat chromosoom 1 een nog ingewikkeldere evolutie heeft doorstaan dan in eerste instantie verwacht werd. De resultaten uit dit hoofdstuk brengen dus nieuwe vragen teweeg, maar ondersteunen tezamen ook een verouderde theorie uit de literatuur waar nu weer nieuw leven in geblazen wordt: namelijk dat het gebalanceerde letale systeem van *Triturus* een uniek geval moet zijn van een plotselinge, evolutionaire 'fout'. En dat deze 'fout' vermoedelijk plaats vond in de voorouder van alle *Triturus* soorten, en in de vorm van een soort 'cytogenetisch ongeluk' (tijdens de vorming van geslachtscellen).

Algemene discussie + Bijlagen

Hoofdstuk 7 van dit proefschrift vat de belangrijkste inzichten samen en bespreekt de bredere impact van het onderzoek als geheel. De evolutie van gebalanceerde letale systemen blijft – ondanks dat er nieuwe inzichten zijn vergaard – nog niet volledig uitgestippeld. Echter, de resultaten van dit proefschrift vormen – na een periode van grofweg veertig jaar zonder veel wetenschappelijke vooruitgang op dit onderwerp – een aanzienlijke stap in het beter begrijpen van de genetische achtergrond van dit fenomeen. Hele genoom studies, in combinatie met RNA studies, zullen in de nabije toekomst nog meer duidelijkheid scheppen. Verder blijken de methoden en technieken die (grotendeels) specifiek voor dit onderzoek zijn ontwikkeld, breed toepasbaar buiten het *Triturus*-vraagstuk. Hier wordt in Hoofdstuk 3 t/m Hoofdstuk 6 dieper op ingegaan (zoals hierboven samengevat), maar in Hoofdstuk 7 wordt hierover verder uitgeweid. In het specifiek wordt aandacht besteed aan de volgende, bredere implicaties van het proefschrift:

1. **Vergelijkbare, genetische fenomenen** – De nieuwe inzichten over gebalanceerde letale systemen kunnen helpen bij het beter begrijpen van andere, soortgelijke genetische fenomenen waarbij heterozygote individuen voordelen hebben ten opzichte van homozygote individuen (bijvoorbeeld de evolutie van sikkelcelanemie in mensen, of '*heterosis*' in planten/gewassen).
2. **Dierwelzijn in de pluimveehouderij** – De in dit onderzoek ontwikkelde genotypering methode 'mxKASP' kan mogelijk bijdragen aan zogenaamde *in ovo* ('nog-in-het-ei') geslachtsbepaling bij (leg)kippen, wat een ethisch alternatief zou kunnen bieden voor het massaal doden van mannelijke kuikens in de pluimvee-industrie. Dat levert momenteel namelijk regelmatig pijn en stress op bij deze dieren – en daarom wordt hard gezocht naar schaalbare en betaalbare oplossingen.
3. **Natuurbescherming en biodiversiteitsonderzoek** – De breed geteste 'NewtCap' technologie kan onder meer worden gebruikt voor het opdoen van evolutionaire kennis, het uitvoeren van genetische monitoring van wilde watersalamander populaties, en het ondersteunen van bijvoorbeeld kweekprogramma's van bedreigde soorten. Dit alles is momenteel van groot belang, zeker gezien de huidige, drastische afname van amfibieën op globale schaal.
4. **Wetenschapscommunicatie en samenwerken** – De betrokkenheid van burgers (ook wel '*citizen scientists*') en andere, gespecialiseerde groepen en organisaties buiten de academie, speelde een zeer belangrijke rol in de totstandkoming van dit proefschrift. Dit onderstreept het belang van interdisciplinaire samenwerking en publieksbetrokkenheid in de wetenschap.

Concluderend biedt dit proefschrift een nieuwe kijk op een van de meest fascinerende genetische paradoxen in de evolutiebiologie: het ontstaan en (voort)bestaan van een erfelijk systeem dat iedere generatie 50% van de nakomelingen opoffert. De combinatie van verschillende toepassingen en benaderingen van de fylogenie, de bioinformatica, en de moleculaire biologie heeft het mogelijk gemaakt om een succesvol onderzoek naar de evolutie van het gebalanceerde letale systeem van *Triturus* uit te voeren en zo (deels) een eeuwenoud raadsel op te lossen. Tegelijkertijd heeft het bepaalde, methodologische ontwikkelingen opgeleverd die qua impact verder (zullen) reiken dan via de toepassingen zoals uiteengezet in dit proefschrift.

© Nienke Prins
(permission to use picture granted)

Appendix II: Awards & Publications

Below is a list of awards and prizes that Manon received, followed by a list of her scientific (peer-reviewed) publications and preprints, a list of her popular (non-peer-reviewed) scientific publications, and a list of the popular science and children's books that feature her (scientific) opinion on certain biodiversity-related matters. The information is provided in chronological order (newest to oldest).

Awards & Prizes

- **Selected as 'Face of Science 2023'** by the Royal Netherlands Academy of Arts and Sciences (KNAW), The Youth Academy (De Jonge Akademie) and NEMO Kennislink, after being nominated by Associate Professor and marine biologist: Dr Lisa Becking.
- **"1st prize for one-minute pitch"** at the Netherlands Society of Evolutionary Biology PhD & Postdoc meeting 2021 (NLSEB2021), for the pitch: "The balanced lethal system in *Triturus*: an evolutionary trap!"
- **"3rd best oral presentation"** at the Programming For Evolutionary Biology Conference 2021 (miniPEB2021), Freie Universität Berlin, for the talk: "Studying the lethality and evolutionary origin of the balanced lethal system in *Triturus* newts".
- **"1st prize for oral presentation"** within subtheme: Global Health - Humans and Animals" at the University of Copenhagen EuroLeague of Life Sciences (ELLS) Student Conference 2017, for the MSc project "The importance of genomics for the conservation management of the critically endangered pygmy hog (*Porcula salvania*)".
- **"Best oral presentation"** at the Benelux International Society of Applied Ethology (ISAE) meeting 2016, for the MSc project "The effect of exposure to visitors on stress in the critically endangered blue-eyed black lemur (*Eulemur flavifrons*) and other primate species at Apenheul Primate Park, the Netherlands."

Scientific publications

- Mars, J., Koster, S., Babik, W., France, J., Kalaentzis, K., Kazilas, C., Martínez-Solano, I., **De Visser, M.C.**, Wielstra, B. (2025). Phylogenomics yields new systematic and taxonomical insights for *Lissotriton* newts, a genus with a strong legacy of introgressive hybridization. *Molecular Phylogenetics and Evolution* 204: 108282
- Kalaentzis, K., Kazilas, C., Arntzen, J.W., Bogaerts, S., Franzen, M., Litvinchuk, S.N., Olgun, K., France, J., **De Visser, M.C.**, Wielstra, B. (2025). Phylogenomics resolves the puzzling phylogeny of banded newts (genus *Ommatotriton*). *Molecular Phylogenetics and Evolution* 203: 108237.

- Koster, S., Polanen, R., Avcı, A., Bogaerts, S., Bozkurt, E., Goudarzi, F., Hemami, M., Olgun, K., Pasmans, F., Steinfartz, S., Üzüm, N.,azan, **De Visser, M.C.**, France, J., Theodoropoulos, A., Wielstra, B. (2024). Discordance between phylogenomic methods due to introgression in Near Eastern mountain newts (*Neurergus*, Salamandridae). *SSRN* preprint 5077292. [Submitted to and Under Review at *Molecular Phylogenetics and Evolution*]

- **De Visser, M.,** France, J., Paulouskaya, O., Brown, T., Fahrbach, M., Van der Ploeg, C., Wielstra, B. (2024) Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying *Triturus*' balanced lethal system, *bioRxiv* preprint 2024.10. 25.620277. [This thesis, Submitted to and Under Review at *Nature Communications*]

- France, J., **De Visser, M.**, Paulouskaya, O., Arntzen, J.W., Babik, W., Cvijanović., M., Ivanović, A., Smith, J., Vučić, T., Wielstra, B. (2024) Genomic evidence suggests the balanced lethal system in *Triturus* newts originated in an instantaneous speciation event, *bioRxiv* preprint 2024.10. 29.620207. [In prep for Journal Submission]

- **De Visser, M.C.**, France, J., McCartney-Melstad, E., Bucciarelli, G.,M., Theodoropoulos, A., Shaffer, H.B., Wielstra, B. (2024) NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts), *bioRxiv* preprint 2024.10. 25.620290. [This thesis, Submitted to and Under Review at *Molecular Ecology Resources*]

- **De Visser, M.,** Van der Ploeg, C., Cvijanović, M., Vučić, T., Theodoropoulos, A., Wielstra, B. (2024) PAV-spotter: using signal cross-correlations to identify Presence/Absence Variation in target capture data, *bioRxiv* preprint 2024.10. 25.620064. [This thesis, Submitted to and Under Review at *Molecular Ecology Resources*]

- **De Visser*, M.C.**, Meilink*, W.R.M., Theodoropoulos, A., Fahrbach, M., Wielstra, B. (2024) Determining zygosity with multiplex Kompetitive Allele-Specific PCR (mxKASP) genotyping *bioRxiv* preprint 2024.10: 25.620256. [This thesis, Submitted to and Under Review at *Molecular Ecology Resources*]
  * = *Authors share a first authorship / equal contribution.*

- Kazilas, C., Dufresnes, C., France, J., Kalaentzis, K., Martínez-Solano, I., **De Visser, M.C**., Anrtzen, J.W., Wielstra, B. (2024) Spatial genetic structure in European marbled newts revealed with target enrichment by sequence capture, *Molecular Phylogenetics and Evolution* 194: 108043.

- Kalaentzis K., Arntzen J.W., Aziz A., Van den Berg V., Beukema W., France J.M., Olgun K., Van Riemsdijk I., Üzüm N., **De Visser M.C.** & Wielstra B.M. (2023) Hybrid zone analysis confirms cryptic species of banded newt and does not support competitive displacement since secondary contact, *Ecology and Evolution* 13(9): e10442.

- Robbemont J., Van Veldhuijzen S., …, **De Visser M.C.** & Wielstra B.M. (2023) An extended mtDNA phylogeography for the alpine newt illuminates the provenance of introduced populations, *Amphibia-Reptilia* 44(3): 347-361.

- Kuijt, M., Oskam, L., Den Boer, I., Dufresnes, C., France, J., Gilbert, M. J., **De Visser, M.C.**, Struijk, R.P.J.H, Wielstra, B. (2023) The introduction of three cryptic tree frog species in the Dutch coastal dunes challenges conservation paradigms. *Amphibia-Reptilia* 44(1), 1-10.
- Koster, S., Prins, N., Dufresnes, C., France, J., **De Visser, M.C.**, Struijk, R.P.J.H., Wielstra, B. (2023) The conservation paradox of an introduced population of a threatened species: spadefoot toads in the coastal dunes of the Netherlands. *Amphibia-Reptilia* 44(1), 11-18.
- De Brouwer, J., Helder, B., France, J., **De Visser, M.C.**, Struijk, R.P.J.H., Wielstra, B. (2023) An isolated crested newt population in Dutch coastal dunes: distribution relict or introduction? *Amphibia-Reptilia* 44(1), 19-26.
- Vliegenthart, C., Van de Vrede, M., Den Boer, I., Gilbert, M., Lemmers, P., France, J., **De Visser, M.C.**, Struijk, R.P.J.H., Wielstra, B. (2023) The limits of mtDNA barcoding for determining the provenance of invasive species: a midwife toad example. *Amphibia-Reptilia* 44(1), 27-33.
- Van der Groot, A., **De Visser, M.**, Hiemstra, A.F. (2022) Smooth newts *Lissotriton vulgaris* observed hibernating in a waterfowl nest. *Herpetological Bulletin*: 162: 41-42.
- **De Visser M.**, Prins E., Bosse M., Crooijmans R., Ter Meulen T. (2022). Maximum longevity and juvenile mortality in zoo-housed mangabeys. *Zoo biology*: 1-11.
- Fahrbach M., **De Visser M.C.,** Wielstra B.M. The hybrid zone between the Italian and Northern crested newts (*Triturus carnifex* and *T. cristatus*) reaches Germany, *Salamandra* 57(3): 428-434.
- **De Visser M.C.,** Liu, L., Bosse, M. (2021). Pygmy hogs, *Current Biology* 31(8):R366-R368.
- Liu, L., Bosse, M., Megens, H.J., **De Visser, M.C.,** Groenen, M.A.M., Madsen, O. (2020) "Genetic consequences of long-term small effective population size in the critically endangered pygmy hog." *Evolutionary Applications* 14:710-720.
- Van Tuijl, C., Van Bochove, K., **De Visser, M.C.** (2019). Genetic structure of badger populations in a fragmented landscape: how do barriers affect populations on a genetic level? *Lutra* 62: 65-76.

Popular Science Articles

- Wielstra, B.M., **De Visser, M.C.**, Stark, T., Struijk, R.P.J.H. (2023) Herkomst van exotische Alpenwatersalamanders blootgelegd, *Schubben & Slijm* 15(3).
- Struijk, R.P.J.H., Prins, N., Koster, S., Putters, N., Jansen, N., Esselaar, J., **De Visser, M.C.**, France, J.M., Wielstra, B.M. (2024) Knoflookpad in Callantsoog: verspreiding, voortplanting en geografische herkomst, *RAVON* 26(1): 2-5.

- **De Visser, M.C.**, Prins, N., France, J.M., Struijk, R., Wielstra, B. (2023) Exotische amfibieën in de duinen ontmaskerd met mtDNA barcoding. *Holland's Duinen* 82(1): 25-29.
- Wielstra, B.M., Den Boer, I., France, J.M., **De Visser, M.C.**, Struijk, R. (2023) MtDNA barcoding van exotische amfibieën in de duinen. *RAVON* 89(2): 26-29.
- Struijk, R.P.J.H., Backx, B., Demirkapu, B., **De Visser, M.**, Wielstra, B. (2022). 'Rise and fall' van een populatie westelijke smaragdhagedissen in Scheveningen. *RAVON* 24(4): 24-27.
- Helder B., De Brouwer J., Ouwehand J., **De Visser M.** & Wielstra B. (2021). Koi-kamsalamander, *Schubben & Slijm* (48): 8.
- Meilink W.R.M., France J., **De Visser M. C.** & Wielstra B. (2021). Balanced lethal systems: an evolutionary mystery. *Frontiers For Young Minds*. 9:632945.
- **De Visser M.C.**, France J.M., Meilink W.R.M. & Wielstra B.M. (2021). Een evolutionair raadsel: het dodelijke chromosoom 1 syndroom in *Triturus*-salamanders, *RAVON* 23(1): 9-12.
- Struijk R.P.J.H, Demirkapu B., **De Visser M.C.** & Wielstra B.M. (2021). Exotische smaragdhage-dissen in de duinen bij Scheveningen, *Kijk op Exoten* (34): 12-13.

Appearance in Popular Science Books

- Future for Nature Academy (2025) *De geheim vissendief - en andere avonturen van jonge natuurhelden*. Details: A fictional children's book featuring various conservation stories inspired by the work of (former) Future For Nature Award winners, including one chapter by **De Visser, M.C.** which focuses on the critically endangered Togo slippery frog in Africa.
- Jansen, J. & Jansen, D. (2022). *PromovendA tot Z – 26 vrouwen in de wetenschap over hun vak en 26 columns van Dolf Jansen over ongeveer hetzelfde.* Details: A popular science book for adults, featuring 'A through Z' stories scientifically backed by young female scientists - including one chapter on climate change (under 'K' for 'Klimaatverandering'), partially written by **De Visser, M.C.**
- Brouwer, D. (2022). *Otje & Ik en de (te) gekke superkracht van mijn axolotl*. Details: A fictional children's book about a pet axolotl with superpowers, featuring a fun and educational 'fact versus fiction' chapter at the end, by **De Visser, M.C.**


Vroege Vogels, BNNVARA, NPO Radio 1
© Anagnostis Theodoropoulos
(permission to use picture granted)

Appendix III: Outreach items & Media attention

Below is a list of outreach items and media attention that relate to Manon's (scientific) work and 'PhD life'. This includes – but is not limited to – projects and interviews that resulted in magazine articles, newspaper articles, radio items, television items, awards, ambassadorships, blogs, vlogs, books, and guest lectures meant for children. The information is provided in chronological order (oldest to newest) and many of these items can still be read, watched, or listened to online today.

- **Blog**: *'Werken met een supercomputer'* [Dutch] / *'Working with supercomputers'* [English]. Bilingual blog explaining how Manon can easily work from home during the pandemic, as she uses a high performance computer cluster. 28-4-2020, **Naturalis website**

- **Interview**: *Naturalis spotlight*. A livestream interview through LiveScience Naturalis (not in person due to the pandemic). Video still available online. 25-6-2020, **LiveScience, Naturalis' YouTube**

- **Video**: '*Manon's DNA talk 1 | een vreemd salamander syndroom*'. A personal vlog explaining my PhD research to a broader audience. 30-6-2020, **Naturalis' YouTube**

- **Video**: '*Manon's DNA talk 2 | wat is DNA?'* A personal vlog explaining what DNA is to a broader audience. 17-9-2020, **Naturalis' YouTube**

- **Video**: *'Manon's DNA talk 3 | Evolutie'*. A personal vlog explaining what evolution is to a broader audience. 25-11-2020, **Naturalis' YouTube**

- **Newspaper Article**: *"Leidse bioloog ontdekt hoe het kan dat de helft van de embryo's van kamsalamanders al in het ei doodgaan [video]".* An article in local newspaper Leidsch Dagblad, describing a *RAVON* publication. *Leidsch Dagblad*. 24-3-2021, **Leidsch Dagblad**

- **Blog**: *'Waarom zijn de eitjes van de kamsalamander gedoemd te sterven?'* [Dutch] / *'Why are crested newt eggs doomed to die?'* [English]. Bilingual blog on Nature Today, explaining the *RAVON* publication. 20-3-2021, **Nature Today**

- **Blog**: *'Grote liefde voor de kleinste varkens'* [Dutch] / *'Big love for the world's smallest pig'* [English]. Bilingual blog about the *Current Biology* 'Quick Guide' published by Manon and co-authors, about pygmy hogs. 29-4-2021, **Nature Today**

- **Radio Interview**: '*Het salamander mysterie'* radio fragment, being interviewed live at the Vroege Vogels studio of BNNVARA, NPO Radio 1. 2-5-2021, **BNNVARA, Vroege Vogels**

- **Interview**: '*Salamander raadsel*' article, belonging to the Vroege Vogels / BNNVARA NPO Radio 1 item. 2-5-2021, **BNNVARA, Vroege Vogels**

- **Video**: '*Manon's DNA talk 4 | Hoe ontstaan soorten*?' A personal vlog explaining the concept of speciation to a broader audience. 7-5-2021, **Naturalis' YouTube**

- **Blog**: Launch of Manon's personal website about science and biology (the 'Wild Science' blog by her company 'Wild DNA'). 10-6-2021, **Wild DNA / Personal site**

- **Blog:** *'Duitse kamsalamanders met "Italiaans" bloed!'*, a personal blog describing a new publication about admixed crested newts in Germany. 18-8-2021, **Wild DNA / Personal site**

- **Video**: '*Manon's DNA talk 5 | DNA barcoding van amfibieën*'. A personal vlog explaining the mtDNA barcoding student projects of our lab. 31-8-2021, **Naturalis' Youtube**

- **Blog**: '*Salamanders and Lizards: 14 ways to tell them apart!*' A personal blog, which Manon worked out during the *scicomm* course of Leiden University's HRM Learning Environment. 23-9-2021, **Wild DNA / Personal site**

- **Lesson/Module**: '*Les 2: een evolutionair raadsel - een dodelijk salamander syndroom*'. An online 'Natuurlab' lesson/module about the balanced lethal system, comprising of assignment, informative texts, and explanatory videos. 1-12-2021, **Naturalis' Natuurlab**

- **Magazine Article**: '*Waarom houden kikkers en padden hun voorpoten voor hun ogen als ze zich bedreigd voelen?'* Q&A section, an interview about why frogs and toads keep their forelimbs in front of their eyes when they are in danger. 1-2-2022, **Quest**

- **Children's lecture**: '*Waarom is de salamander een superdier?'* A lecture designed for children, about the fascinating traits of salamanders, provided at the Naturalis' 'tribune' room. 13-2-2022, **MuseumJeugdUniversiteit**

- **Blog**: '*Monkey Business: publicatie in '*Zoo Biology'', a personal blog about one of Manon's papers that came out and focused on population demography of captive mangabeys in zoos. 6-4-2022, **Wild DNA / Personal site**

- **Blog**: '*De pygmy hog: nu te zien bij 'favo items' in LiveScience!'* A personal blog about Manon's efforts to bring the pygmy hog specimen to the public temporarily, at the LiveScience hall of the Naturalis' museum. 3-5-2022, **Wild DNA / Personal site**

- **Blog**: '*One pond, four species!'* A personal blog about field work during which Manon observed all native, Dutch newt species at once in the same pond close to her home – a rare sight. 8-5-2022, **Wild DNA / Personal site**

- **Magazine Article**: *'Deze salamander doet aan parachutespringen',* an article by *KIJK Magazine* featuring Manon's opinion on a 'skydiving' salamander study. 25-5-2022, **KIJK Magazine**

- **Magazine Article**: 'The tragedy of *Triturus*', an article in Leiden Science Magazine, written by a first year BSc student in response to the guest lecture Manon provided at the Symposium 'Populair Wetenschappelijk Schrijven'. 9-6-2022, **Leiden Science Magazine**

- **Video**: *'Bioloog Manon de Visser over mangabeys en dierentuinen'.* A personal video/interview about a mangabey paper (a journal publication in *Zoo biology*, with Leiden University as one of the affiliations). In the video Manon discusses the importance of collaborations between zoos, museums and universities. 12-6-2022, **Naturalis' YouTube**

- **Magazine Article**: *'Kikker kan door kleine oren niet goed springen*', an article by *KIJK Magazine* featuring Manon's opinion on a recently published paper about 'clumsy mini-frogs'. 17-6-2022, **KIJK Magazine**

- **Keynote Talk**: *'I see the beauty in DNA and the 'mistakes' that nature makes therein*', a keynote lecture at the Hooglandse Kerk during Leiden2022 (the Leiden European City of Science 2022), for the art exhibition 'TRINITY: contemporary triptych in Leiden'. 28-6-2022, **Leiden2022 / TRINITY / Wild DNA**

- **Children's lecture**: A guest lecture, designed for children, provided by Manon to kick-start the 2022 Summerschool Junior of Leiden University (by *ICLON*). Both children and their parents were present. Manon discussed what it is like to be a biologist, and talked about salamanders. 25-8-2022, **University/college**

- **Video**: '*Manon's DNA talk 6 | Wat zijn hybriden?'* A personal vlog explaining the nature of hybrid species, using crested vs. marbled newts and hybrids as an example, showcasing amazing fieldwork footage. 6-10-2022, **Naturalis' YouTube**

- **Podcast Item**: At the *Podcast NPO Radio 1: "Alledaagse Vragen"* of BNNVARA/NPO Radio 1, Manon was interviewed about the basic biology/genetics behind (human) sex ratios. 26-10-2022, **BNNVARA, "Alledaagse Vragen"**

- **Media Storm / Book**: For a popular science book featuring young, female scientists, Manon was asked to provide input. The book is called '*PromovendA tot Z'*, and includes Manon under Chapter K (which stands for 'Klimaatverandering' / 'Climate change'). 3-10-2022, The book got discussed by; **AD, NPO Radio 1, Eva Jinek, Parool, New Scientist, De Volkskrant & more**.

- **Book**: For a children's book about a boy and his pet axolotl, Manon provided the end chapter. The book is called *'Otje & ik'* and tells a fictional story. In her chapter, Manon explains in a fun way what is salamander fact, and what is salamander fiction ('feit of fabel'). 16-11-2022, see **Personal site / Wild DNA**

- **Blog**: '"*Aliens in de duinen!?" Exotische amfibieën ontdekt'.* A personal blog describing the articles that got pre-published online (in the journal *Amphibia Reptilia*) about exotic amphibians in the Dutch dunes. 1-12-2022, **Wild DNA / Personal site**

- **Blog**: *'Boomkikkers ontmaskerd'* [Dutch] / *'Tree frogs exposed'* [English]. Bilingual blog on *Nature Today*, explaining the *Amphibia Reptilia* papers. 1-12-2022, **Nature Today**

- **Video**: '*Manon's DNA talk 7 | "Buitenlandse boomkikkers"'* A personal vlog explaining the results of the amphibian mtDNA barcoding studies of her research group, including amazing field- and lab-work footage. 1-12-2022, **Naturalis' YouTube**

- **Media Storm / Newspaper Articles**: Different media picked up on the *Amphibia Reptilia* publications regarding exotic amphibians in the Dutch dunes and reported the slightly shocking results. 12-2022, **NOS, AD, ATLAS, Omroep West, Noordhollands Dagblad, West Wordt Wakker (Radio Interview), Scientias, Kidsweek & more**.

- **Blog:** 'Salamanders overwinteren in vogelnesten', a blog about a discovery in the Leiden canals where smooth newts were observed hibernating in a fowl's nest, that Manon described with co-workers as a 'Natural History Note' in *The Herpetological Bulletin*. The item incorporated Manon's commentary on the matter. 26-12-2022, **Nature Today**

- **Blog**: '*The chimp and the bananas: how I helped design the new Evolution exhibition of Naturalis*'. The exhibition designers of the Naturalis museum asked Manon for help designing their concept and content for in the new museum hall called 'Evolution'. Manon wrote a blog about how her ideas got incorporated. 14-1-2023, **Wild DNA / Personal site**

- **Magazine Article**: *'Kippen-DNA vervuilt genen wilde oerkip',* KIJK Magazine asked Manon to comment on the dangers of admixture/hybridization between domestic and wild chicken, related to a paper published by the National University of Singapore. 20-1-2023, **KIJK Magazine**

- **Newspaper Article**: *'Ondergronds oeramfibie duikt op'*; a Bionieuws article about a paper by paleobiologists describing ancient caecilian fossils, featuring Manon's comments on it. 27-1-2023, **Bionieuws**

- **Radio Interview:** *SLAM!* interviewed Manon on live radio, talking about dodo- dinosaur- mammoth- and salamander DNA (and the pros and cons of de-extinction). 2-2-2023, **SLAM!**

- **Interview / Newspaper Article**: *'Bioloog Manon de Visser heeft 'een voorliefde voor slijmerige beesten''.* NRC published an extensive, personal interview with photos taken at Naturalis/LiveScience to shine a light on what it is like to be a young biologist and PhD candidate. 3-3-2023, **NRC rubriek 'Jong Geleerd'**

- **Award:** Manon was selected as *'Face of Science 2023'* by the KNAW, the Youth Academy (Jonge Akademie), NEMO Kennislink, and Stichting Lira. *Faces of Science* is a KNAW initiative that gives a platform to successful, young academics so they can openly talk/blog/vlog about anything that inspires them and their work. 9-3-2023, **KNAW / NEMO / Faces of Science**

- **Newspaper Article**: *'Waarschuwingskleuren evolueren stapsgewijs',* a Bionieuws article about a study related to aposematism / camouflage evolution in amphibians, featuring Manon's comments. 24-3-2023, **Bionieuws**

- **Newspaper Article**: *'Er zijn steeds minder padden, maar vrijwiligers proberen het tij te keren'*, Manon was interviewed to talk about the big toad migration in spring, and about how toads benefit from volunteers that help them cross roads. 27-3-2023, **De Volkskrant**

- **Radio Interview**: *'Kokerjuffers, de Nationale Bloem en bonttest'* A reporter from Vroege Vogels joined Manon's students and herself during amphibian-related fieldwork. The talks, and muddy sounds, were recorded and broadcasted. 16-4-2023, **BNNVARA, Vroege Vogels**

- **NEMO Blog**: *'Amfibieën vertellen ons hoe gezond de natuur is'*, as Face of Science and NEMO Kennislink blogger, Manon writes about topics that inspire her. English versions are on her personal site. 25-4-2023, **KNAW / NEMO / Faces of Science**

- **NEMO Blog**: *'Zes redenen waarom amfibieën wereldwijd achteruit gaan'*, as Face of Science and NEMO Kennislink blogger Manon writes about topics that inspire her. English versions are on her personal site. 25-4-2023, **KNAW / NEMO / Faces of Science**

- **Interview**: *'Evolutie is gedachteloos, lui en imperfect - en dat is precies waarom het zo goed werkt!'* The personal Faces of Science profile of Manon got launched, portraying a lengthy interview and links to her KNAW / NEMO / Faces of Science blogs. 1-5-2023, **KNAW / NEMO / Faces of Science**

- **Video**: *'Welke afwijkingen hebben salamanders in het DNA?'* A YouTube video by KNAW presenting Manon and her research, which was recorded at Naturalis Biodiversity Center. 1-5-2023, **KNAW YouTube**

- **NEMO Blog**: '*Wat is jouw lievelingsdier*? En andere gekmakende vragen!', as Face of Science and NEMO Kennislink blogger Manon writes about topics that inspire her. English versions are on her personal website. 28-6-2023, **KNAW / NEMO / Faces of Science**

- **Magazine Article**: *'Deze zeeslang kan weer kleuren zien'*, KIJK Magazine asked Manon to comment on a new paper about the evolution of color vision in sea snakes and incorporated her commentary in their article. 13-7-2023, **KIJK Magazine**

- **Blog**: '*Exotische Alpenwatersalamanders blijken wijdverspreid'* A Leiden University press release, coinciding with a Nature Today blog, about a new paper focused on identifying populations of exotic alpine newts across The Netherlands. 2-8-2023, **Nature Today & Leiden University**

- **Radio Interview**: *'Evolutionair foutje? Het mysterie rondom salamandereieren'* A live, radio interview about Manon's PhD project an Halloween hobby. 4-9-2023, **NPO Radio 1, "Villa VdB"**

- **Video**: '*Manon's DNA talk 8 | "Bandsalamanders zijn als olie en water"'* A personal vlog explaining the results of a study of the group on banded newt hybridization. 5-9-2023, **Naturalis' YouTube**

- **Podcast item**: *'#6 – Staan voor de zaak: Henk Grol, verslaafd aan zitten, Golden Earring ontleed en een fout in de evolutie'*, the Villa VdB radio interview was selected as one of the best four interviews of the week and therefore ended up in an additional Villa VdB podcast episode. 8-9-2023, **Podcast "Villa VdB"**

- **Magazine Article**: *'Rondworm kweekt enorme mond om familie op te eten'*, *KIJK Magazine* interviewed Manon on the topic of cannibalism, as a response to a recent paper on the genetics and cannibalistic traits of certain roundworm species. 15-9-2023, **KIJK Magazine**

- **Interview**: *'Kikkerbillen: weten we eigenlijk wel wat we eten?'* Together with *KRO-NCRV* and the *Keuringsdienst van Waarde* Manon co-supervised two BSc students to analyze the DNA of frozen frog legs from the supermarket, which led to a shocking reveal. 20-9-2023, **KRO-NCRV & Keuringsdienst van Waarde**

- **Blog**: *"Kikkerbillen eten? Dit moet je eerst weten…"* In response to the shocking reveal of the *KRO-NCRV* and *Keuringsdienst van Waarde* frog legs test, Manon wrote a personal blog, which got picked up by *Faces of Science* / NEMO Kennislink. 20-9-2023, **Wild DNA / Personal site & KNAW / NEMO / Faces of Science**

- **Interview / Magazine Article**: *"Leuk stel: Manon de Visser en Chris van der Ploeg uit Breda trekken regelmatig samen de nacht in"* Journalists from the *ANWB Kampioen Magazine* interviewed both my husband and me to talk about our passion for 'night walks in nature'. 28-9-2023, **ANWB Kampioen**

- **Magazine Article**: *'Vormen kruisingen tussen verschillende diersoorten een probleem?'*, *KIJK Magazine* interviewed me regarding a recent discovery of a hybrid offspring of a dog and fox and the genetics and dangers tied to such phenomena. 9-11-2023, **KIJK Magazine**

- **Lesson/Module**: *'NEMO Science Night 2023: Getsie, gatsie, genetica',* as *Face of Science* I was invited to present during the *NEMO Science Night 2023*, to provide biology teachers with materials and inspiration that would support their DNA/genetics classes. I uploaded many of these materials on my website shortly after, too. 23-11-2023, **Wild DNA / Personal site & KNAW / NEMO / Faces of Science**

- **Blog**: *'Wat was er eerst, de kip of het ei?'*, a journalist from *NEMO Kennislink* reached out to discuss what came first: the chicken or the egg? And my comments were incorporated in an online article. 8-12-2023, **NEMO Kennislink**

- **Blog**: *'Op bezoek bij het Douane DNA lab'*, together with two students of our group I visited the DNA laboratory of the Dutch customs in Amsterdam. I wrote a personal blog about. 22-1-2024, **Wild DNA / Personal site**

- **Lecture / Workshop / Theatre**: *'Moving Animals Festival Naturalis: Workshop Exoten, van welkom tot invasief'*, together with a historian from *Maastricht University*, I organized a workshop for the broader public about exotic species at *Naturalis* during the *Moving Animal Festival* – including a small acting performance. 2-2-2024, **Naturalis Biodiversity Center**

- **Magazine Article**: *'We danken ons supersnelle zenuwstelsel deels aan virussen',* KIJK Magazine requested me to comment on a new paper that described the evolution of the neural system, and on how viruses played a role in it. 15-2-2024, **KIJK Magazine**

- **Blog**: *'Mijn stokoude dieren-top10',* as Face of Science and NEMO Kennislink blogger Manon writes about topics that inspire her. English versions are on her personal website, 19-2-2024 **KNAW / NEMO / Faces of Science**

- **Blog**: *'Wat is taxidermie?'* A personal blog about taxidermy, as Manon started working as a taxidermy volunteer for the Naturalis team in February, 2024. 21-3-2024, **Wild DNA / Personal site**

- **Blog**: *'Wilde wezens juichen om een bosbes van golfbal-formaat'* A personal blog about how the natural instinct of humans and other animals to be fascinated by objects that are out of the ordinary. 25-3-2024, **Wild DNA / Personal site**

- **Blog**: '*Alweer een jaar 'Face of Science' winnaar bij de KNAW'* A personal blog about being a Face of Science. It was picked up and published by NEMO too. 28-3-2024. **Wild DNA / Personal site & KNAW / NEMO / Faces of Science**

- **Ambassadorship**: an ambassadorship for *'De Week van de Biologie 2024'* events, as selected by NIBI (the Dutch Institute for Biology). 1-4-2024, **NIBI**

- **Magazine Article**: *'Hoelang fokken mensen al kippen voor eieren?'* KIJK Magazine asked Manon to comment on a new paper about the domestication of chickens. 3-4-2024, **KIJK Magazine**

- **Blog / Ambassadorship**: '*Biologie zit in mijn DNA - en ook in dat van jou'* A personal blog, written for the NIBI Ambassadorship, about 'De week van de Biologie 2024'. 17-4-2024, **Wild DNA / Personal site**

- **Blog**: *'Minder biologie in onderwijs en wetenschap? Niet als het aan Wild DNA ligt…'*, A personal (/slightly political) blog about how primary school and high school level biology education is threatened – and what she does about it through Wild DNA jobs. 13-6-2024, **Wild DNA / Personal site**

- **Television Interview**: *'Dierbeschermers: niet-schattige dieren krijgen te weinig aandacht',* a national television item on a news channel for children featuring Manon standing up for 'less popular' animals, such as amphibians and reptiles. 27-6-2024, **NOS Jeugdjournaal / NPO Zapp**

- **Newspaper Article***: 'Kleuren en geld gaan voor: meest bedreigde vissoorten krijgen de minste aandacht'*, a article by NOS, featuring the commentary of Manon after being interviewed, regarding the so-called conservation bias in relation to a new reef fish study. 17-7-2024, **NOS**

- **Magazine Article**: '*Deze vis 'loopt' op de zeebodem en 'proeft' zijn prooien*', KIJK Magazine requested Manon's commentary on a new paper about the evolution of 'magical' appendices in gurnard fish, and included it in their article. 27-9-2024, **KIJK Magazine**

- **Blog**: *'Borneo avonturen: congres en rondreis!'*, a personal blog about the scientific conference in Borneo in 2024, and the travelling adventures connected to that event. 9-10-2024, **Wild DNA / Personal site**

- **Newspaper Article**: *'Zeven nieuwe kikkersoorten ontdekt die niet kwaken, maar fluiten'*, an item about the discovery of 'whistling' frogs, including Manon's thoughts on the discovery. 15-10-2024, **De Volkskrant**

- **Radio Interview**: '*Bioloog Manon de Visser over vreemde gedragingen en geluiden bij dieren*', a live radio item, discussing the odd behaviors observed in the animal kingdom. 20-10-2024, **NPO Radio 2 "Vroeg Pieken!"**

- **Podcast Item:** *'Can animals be childfree? (with Dutch Geneticist Manon de Visser)'*, a podcast interview published on Spotify (sneak preview), discussing the global news surrounding a Dutch, mass sperm donor, and general, evolutionary, reproductive strategies in nature/wildlife (+ the animal 'childfree lifestyle'). 24-12-2024, **Dinky Podcast (USA)**

- **Newspaper article**: *'Dit zijn de vijf bijzonderste nieuwe diersoorten van 2024'*, a newspaper from Belgium incorporating Manon's thoughts from the article by De Volkskrant that was published weeks earlier and focused on the whistling frogs discovery. 30-12-2024, **De Morgen (Belgium)**

A guiding light

© Marijn Scheeres
(permission to use picture granted)

Appendix IV: Acknowledgements / Dankwoord

Completing a PhD is not an easy-peasy-lemon-squeezy thing to do. Yet, I have not been able to find a dissertation in which the author took the opportunity to thank themselves. But I believe self-recognition matters and that it should be less frowned upon. Therefore, I hereby – briefly, but wholeheartedly – thank myself for having done an amazing job. That being said, I could have never, ever conducted this research without the support of so many incredibly sweet and talented people – likely including you! So, the rest of this appendix is dedicated to you all. I will not mention names (except those of my cats, but that is because I think they are actual saints that require my worshipping or else bad things happen). Still, I am confident you will know what applies to you.

Colleagues

First of all, I am beyond grateful to my supervisors, mentors, colleagues, and students – especially those from IBL/Leiden University and Naturalis Biodiversity Center – for believing in me and in this project. Thank you for making this journey memorable and enriching.

I will never forget the fun and messy field work in the ponds, the inspiring teaching moments, the thrilling media performances, the crazy lab-disasters, the COVID-era Zoom meetings (and all the cats I could digitally worship thanks to those), the lab visits abroad, the fruitful debates, the countless shared meals and celebrations… And a special thanks goes to all members of my PhD monitoring-, dissertation-, and opposition committees.

I also deeply appreciate all my co-workers from beyond IBL and Naturalis, including those from CML, the Hortus, RAVON, KNAW, NEMO Kennislink, and different Universities, government institutes, museums, and zoos in The Netherlands. And the same goes for the many international collaborators. Amazingly, you entered my life and stayed connected from all corners of the world; Germany, England, France, Poland, Serbia, Greece, the United States, the Emirates, and beyond. Thank you so much for your invaluable contributions!

Friends

To my friends – whether research-related or simply there for mental support – thank you for your advice, your encouragement, and your sparkle.

Leiden friends from IBL, Naturalis, CML, Hortus, LUMC: I want to say I admire you so much! I'll never forget the 'Among Us' betrayals that somehow kept us sane during

lockdowns, the extravagant Christmas parties, the awkward online fitness sessions, the dinners, the home-baked-office-cookies (sorry for never returning the favor!), the creative painting and crafting times, the taxidermy lessons, and the joy of collaborating with you – especially in LiveScience. Thanks for the laughs and the co-venting. You made my 'off-work' moments in Leiden fun and exciting!

Childhood, college, scare-acting, and body-art friends: I deeply appreciate you all. You always help me through the rough times (even if you are not aware of it). And those who engage with me or my *scicomm*-work primarily through social media or my WildDNA-website: know that it means a lot to me that we became, or have simply always been, remote/online friends. Many of you are just one text or tag away – and I value that.

Family & Partner

If you are part of my family – whether by blood or 'aangetrouwd' – you already know I love you and am filled with gratitude for your support. Shout out in particular to the 'ChriRiLauManSte', 'Fries Before Guys', and 'Ploegies' *gang*s!

Foremostly, I want to thank my parents. Pap, mam: thank you, for teaching me to live the way I want (not the way others expect) and for raising me with the '010' and '*gewoon-doen-joh*' mentality. I would have not been where I am today if it were not for your parental love.

Also, I want to deeply thank my sister and her spouse. I enjoy going to concerts and events with you guys, and discussing the good and the bad stuff of life in our homes, and sending funny reels to each other – over and over. Sis, thank you for being just that: my sis, as well as my fellow 'environmental expert', busy bee, and even paranymph!

And hubby: I do not know how you put up with me, honestly… But you are, somehow, always there. You are my family, my best friend, my partner in crime, and even my anti-fireworks comrade, handyman, mind reader, paranymph, co-author… (*do you secretly have superpowers?!)* You held me close – through my highs and lows – despite your own PhD project, marathons, and other worries. I love you so, so much. *감사합니다 - "gam-sa-ham-ni-da"*

Cats

As foretold, I finally have to glorify Siri and Goya. I thank them for being so adorably cute (and for being blissfully unaware of it). They – like the cats that already came before them – remind me that cats are our true saints and saviors. Take it from a crazy cat lady, that one that became quite fond of newts, too.


Pixabay

Appendix V: *Curriculum Vitae*


© Frank Boots*

Manon de Visser was born on June 11, 1993, in Spijkenisse, The Netherlands. After completing high school at Emmauscollege Rotterdam in 2011, she pursued a BSc in Biology at Utrecht University, focusing on animal behavior and cognition. She concluded her BSc with a research project on nestling sea turtles in Costa Rica in 2014 and graduated in 2015.

That year, she enrolled in a 'double MSc' specialization in Biology at Wageningen University & Research, where she studied a wide range of animals (incl. primates and hogs) while closely collaborating with zoos and other NGOs. Her MSc research covered different fields, such animal genomics, conservation, ecology, behavior, and population management.

After finalizing her MSc in 2018, Manon worked as a project manager in ecology at Econsultancy and simultaneously as an environmental DNA (e-DNA) researcher and science communication officer at Datura Molecular Solutions B.V. This included plenty of desk work, as well as field work, that related to the legal protection of Dutch flora and fauna (incl. birds, reptiles and mammals such as bats and badgers). She worked for these companies until late 2019. Then, she joined the Institute of Biology Leiden at Leiden University, as well as Naturalis Biodiversity Center, to begin her PhD at the Wielstra Lab, studying the evolution of balanced lethal systems in *Triturus* salamanders. This research was fully funded by the European Research Council.

Alongside her PhD project, Manon also contributed to other herpetological studies and she actively worked towards attaining broader public outreach. In line with this, Manon also founded Wild DNA in 2022: her personal science communication company, in which she is self-employed. Following a formal contract extension (through a COVID-19 compensation arrangement), Manon continued her work at the Wielstra Lab as a postdoc (or 'pre-doc') in 2024, again funded by the European Research Council, in order to conclude her *Triturus* studies.

As of April 2025, Manon will continue her career in zoo science, as she will join the Conservation & Science Center of the zoo in Rotterdam: Diergaarde Blijdorp. She will work there as a Research Officer, overseeing scientific projects at the zoo, while also remaining affiliated with Naturalis Biodiversity Center.

**(permission to use picture granted)*

Diergaarde Blijdorp (Rotterdam Zoo)

Appendix VI: Career story

Childhood

Manon de Visser was born in Spijkenisse, The Netherlands, on June 11, 1993. As a naturally curious child, she developed a love for both drawing and animals, and frequently visited petting farms and zoos. Growing up in Rotterdam with her parents and sister, she visited Rotterdam Zoo (Diergaarde Blijdorp) most often – at times weekly – and this would shape her future career aspirations.

During her school years at Ds. J.J. Buskesschool (primary school) and Emmauscollege Rotterdam (high school), her favorite subjects were art and biology, and she aspired to be a zoologist. In 2011, before graduating high school, she was invited to deliver the final graduation speech at the New Luxor Theatre in Rotterdam, representing all HAVO, VWO, and Gymnasium-level pupils that graduated that year. Despite her social anxiety, she accepted the challenge and spoke in front of a large crowd.

Bachelor

In 2011, Manon initiated her BSc in Biology at Utrecht University (UU). She joined the editorial team of BioSCOPE magazine, worked as a Teaching Assistant for a BSc-level animal anatomy course, and worked as a communications intern for Stichting Trésor. She pursued a Minor comprising veterinary sciences, as well as psychology related courses, and she conducted her BSc thesis research on the influence of global warming on temperature-dependent sex determination in sea turtles. For this research, she spent time at Reserva Playa Tortuga in Costa Rica, where she also assisted in caiman and mammal studies and road kill studies (by conducting dead animal street monitoring). During her Bachelor studies, Manon received her 'Safe Microbial Techniques' certification (2011), and also followed additional evening workshops offered by UU in 2014 ('Career business for Biologists', 'Herpetology', and 'Bee keeping for Biologists'). Moreover, during her BSc she volunteered at a Dutch hedgehog rehabilitation centre (EgelopvangZHZ). She graduated her Bachelor's program in 2015 with a final GPA of 3.95.

Master

In 2015, Manon began her MSc in Biology at Wageningen University & Research (WUR), specializing twice by undertaking two Master's thesis projects and two Master's

internships (instead of just one of each). She conducted her studies at the Resource Ecology Group and the Animal Breeding and Genomics Centre. Her favorite project involved whole-genome data analysis of pigs, focusing particularly on the pygmy hog – the smallest and most threatened pig species in the world – in collaboration with the Pygmy Hog Conservation Programme in Assam, India, and Durrell Wildlife Conservation Trust at Jersey Zoo. During this time, she also earned a certificate in Basic Breeding Programme Management through a European Association of Zoos and Aquaria (EAZA) course at ARTIS, and the 'Companion Animals Advanced' certificate via PTC+ (a practical training centre for agriculture, horticulture and animal husbandry).

The rest of her MSc projects included a study on blue-eyed black lemur stress-related behaviors and -hormones (at Apenheul Primate Park and the Biomedical Primate Research Centre), a study on the demography of captive grey-cheeked and black-crested mangabey populations (at GaiaZOO), and a study on chimpanzee welfare and stress-behaviors via ChimpanZOO (a project of the Jane Goodall Institute Global, conducted at Apemanagement). For the latter project, Dr. Jane Goodall personally acknowledged Manon's contributions via handwritten letters. Manon also won several awards for oral presentations at international conferences and – in line with this – became a Teaching Assistant for a WUR course on oral presentation skills. She also participated in an Academic Consultancy Traineeship with BirdLife Netherlands, served on the program's 'Master's Committee', and was recruited as student member of the Appointment Advisory Committee (BAC) for a tenure-track selection procedure at the Animal Breeding and Genomics Centre. She also obtained her certification in Laboratory Animal Science (Article 9 of the Experiments on Animals Act law), allowing her to more easily conduct certain animal studies in the future. During her studies, Manon had always worked several side-jobs (in sales, entertainment, retail, etc.) to support her college fees. She eventually graduated her Master's program in 2018, with a final GPA of 4.0.

Start of career

Right after obtaining her MSc diploma, Manon sought out a suited position in the zoo community – but found opportunities elsewhere. She worked at Datura, a molecular research agency specializing in environmental DNA monitoring, and at Econsultancy, an ecological advisory office focused on the legal protection of Dutch biodiversity, in 2018 and 2019. She gained extensive field and desk experience, and worked round the clock to research Dutch flora and fauna. In this period, she obtained her VCA certification ('Veiligheid voor Operationeel Leidinggevenden'). Meanwhile, she kept raising public awareness for the biodiversity and climate change crises in her free time (something she already did as a BSc and MSc student), for instance by raising funds for Stichting

Herpetofauna through modelling at body art events, or by volunteering at other NGOs like the Future For Nature Academy (as an events organizer), Rotterdam Zoo (as an educator), the Jane Goodall Instituut NL (as a social media manager), and IVN Natuurontdekkers (as a mentor at summer camps for children).

PhD and 'pre-doc'

In 2019, Manon started her PhD at the Wielstra Lab at Leiden University (LU, Faculty of Science), researching the balanced lethal system in *Triturus* newts. Her work was conducted at the Institute of Biology Leiden (IBL, Animal Sciences) in collaboration with the Understanding Evolution research group at Naturalis Biodiversity Center, and it was funded by a European Research Council (ERC) Starting Grant awarded to Dr. Ben Wielstra. Next to conducting her LU research, she concluded in-depth scientific courses such as via ForBio ('Target Enrichment', 2020) and Physalia ('Phylogenetic comparative methods' and 'Phylogenomics', 2021), as well as other types of education, like trainings on *soft skills*, scientific conduct, and project management via LU's HRM Learning & Development (2020-2022) and the 'Summer School Science Communication' via the LU research group Science Communication and Society (2024). She also taught several university-level courses herself, mostly the yearly Evolutionary Biology 2 LU Bachelor course (2020-2024) and the yearly Genomic Architecture' LU Master course (2022-2024). She also served as an events organizer for the Naturalis PhD Council (from 2020 until 2022) and she joined the JongerenXNaturalis youth initiative as a junior 'core group' expert for one year (in 2023).

Through her PhD and 'pre-doc' experiences, Manon deepened her understanding of evolution, DNA, biodiversity, and the urgent need to protect amphibians worldwide. She supported with the design of some of the educational texts that are on display in Naturalis' permanent 'Evolution' exhibit, she developed a high school level, online module about her PhD research topic through Naturalis' Natuurlab environment, and she created several, educational vlogs for Naturalis' YouTube channel dubbed 'Manon's DNA talks'. Also, she presented some of her findings at scientific conferences, such as at the 'mini PEB 2021' conference by the Freie Universität Berlin (winning the 'third best oral presentation' award), the Dutch 'NLSEB 2021' PhD & Postdoc meeting (winning the 'first prize for a one-minute pitch' award), and the Dutch 'NWO Life 2024' conference (with a two-day full display of her scientific art, due to being one of ten 'Visualization Contest' nominees). However, most notably was Manon presenting her work orally at the largest and the openly feministic 'World Congress of Herpetology' in 2024 in Borneo. Furthermore, Manon started volunteering for the Reptile, Amphibian & Fish Conservation Netherlands (RAVON) in her free time, by monitoring amphibians in the forests around her home town Breda and simultaneously used her monitoring permits to teach her students certain fieldwork skills.

Science Communication

In 2022, Manon founded her own science communication and arts business called 'Wild DNA'. Through Wild DNA, she co-authors popular science books and children's stories, delivers keynote talks and guest lectures, fact-checks biology-related information, and creates DNA art (primarily jewelry featuring real, extracted DNA strands). Her efforts in science communication led to her recognition as a 'Face of Science' by the Royal Netherlands Academy of Arts and Sciences (KNAW), The Young Academy (De Jonge Akademie), NEMO Kennislink, and Stichting Lira in 2023. Her media appearances include items by NRC, AD, NU.nl, NPO, Bionieuws, SLAM!, NOS Jeugdjournaal, De Volkskrant, and more, and she also was appointed as one of the ambassadors of the Dutch 'Week van de Biologie' by the Netherlands Institute for Biology (NIBI) in 2024.

Since 2024, Manon has additionally volunteered at Naturalis Biodiversity Center by supporting the taxidermy team with preparing study skins and skeletons of mostly small birds and mammals, often during live, public demonstrations at the LiveScience hall of the museum. She also regularly helps maintain and clean the specimen on display at the museum, including fragile, dinosaur skeletons. Also, as being a biologist as well as a vocal conservationist can be mentally demanding, Manon has been working as a 'scare-actress' at Walibi Holland's yearly *Halloween Fright Nights* events since 2017. While she sees this as a fun way to blow off steam, it also further enhances her knowledge and skills in engaging with the general public via another sector of the leisure industry: theme parks. This uniquely complements her experience working in zoos and museums.

Future prospects

Looking ahead, Manon plans to stay committed to her professional career in zoology, conservation (genomics), and science communication after being awarded the degree of Doctor. Her goal is to build effective collaborations, test new ideas, and bring the public closer to biodiversity science than ever before – all in the hope of supporting the ongoing study and protection of threatened species, including *Homo sapiens*, worldwide.

Recently, Manon took on a position as *Research Officer* at the Conservation & Science Center of Rotterdam Zoo: the place that is full of her childhood memories, and that sparked her dream of becoming zoologist already at a very young age. She thanks the pygmy hippo of the zoo, Eveline, specifically for inspiring the research ideas that contributed to her being eventually selected for the role. She will also remain affiliated with Naturalis Biodiversity Center and aims to strengthen the scientific collaborations between the two knowledge institutes. She will start working at the zoo in April, 2025.

Triturus *eggs wrapped in leaves of aquatic plants*