Departmental Coversheet

UNIVERSITY OF OXFORD

BIOLOGICAL SCIENCES

GRANT PROPOSAL SUBMISSION

Candidate Number	**1033961**		
Title	**Comparative study of the potential role of transposable elements in vertebrate regeneration**		
Project duration	**3 years**	**Total amount requested for project expenses**	**£222,717.00**
Abstract (Maximum 200 words)	In their evolutionary history, transposable elements (TEs) have been selected on their replicative ability and exapted by their hosts due to their gene-regulating ability. The association between TE activity and stem cell identity through this exaptation has led to the speculation that the TEs' gene regulatory function might regulate animal regeneration. The work proposed searches for and annotates putative regeneration-related TEs in vertebrate genomes using criteria such as proximity to regeneration-related genes and upregulation in regenerating tissue. It then seeks to establish causal relationships between these TEs and regenerative function in model vertebrates by manipulating target TE activity in wet-lab experiments. If causal relationships can be established, it further aims to compare the involvement of TEs in the regeneration of different vertebrates. This study involves both bioinformatic approaches that compare the TE profile surrounding regeneration-related genes among genomes and wet-lab experiments investigating whether regeneration is affected when specific regeneration-related TEs are silenced. This study addresses a previously uncharacterised relationship; it helps understand the genetic control of regenerative function in different vertebrates by tracing the evolutionary history of regeneration-related TEs. It also harbours potential means through which regenerative function can be regulated, contributing to approaches to stimulate regeneration in non-regenerating animals.		

DO NOT ENTER YOUR NAME ANYWHERE IN THIS DOCUMENT.
DO NOT ENTER YOUR CANDIDATE NUMBER ANYWHERE ELSE.

INTRODUCTION

Transposable elements (TEs) are a substantial component of eukaryotic genomes[1]. They are mobile genomic elements that frequently move around the genome by different mechanisms, such as 'cut-and-paste' and 'copy-and-paste' mechanisms[1]. They can also amplify within the host genome independent from host replication mechanisms, resulting in their high prevalence and widespread distribution across the genome[1]. They are further divided into subgroups by the different replicating mechanisms they use and the evolutionary origins they share[2]. They play significant roles in host gene expression by providing regulatory elements as well as coding and non-coding RNAs[1].

Previous studies have established the association of TEs with stem cell identity in mammals: multiple but very different TE families are active in the early embryonic development of the human and the mouse, during which stem cells are not yet differentiated[3,4]. For example, a family of TEs called LINE1 acts as a nuclear RNA scaffold that directs gene expression programmes critical for the maintenance of stem cell identity in the mouse[5]. However, little has been studied on the relationship between TEs and vertebrate regeneration, an ability that differs dramatically among vertebrate species[4]. Since stem cells are critical for regeneration, the widespread role of TEs in host gene expression and stem cell identity control have made TEs potential candidates that might play a role in the regulation of regenerative ability in the vertebrates.

TEs possess host-independent replication mechanisms and act selfishly to proliferate themselves. They are related to viruses and endogenous viral elements (EVEs), another type of mobile genomic element, in that they are similar in the mechanism they use to amplify within host genomes[6]. EVEs are DNA sequences derived from the chromosomal integration of viral DNA into the host germ cells[7]. The endogenisation events of ancient viruses and their subsequent amplification events are found throughout genomes of all eukaryotic lineages, contributing to the widespread nature of these elements[7]. EVEs are not entirely useless genomic relics; multiple EVE lineages – such as the HERVH family in humans and MERVL family in mice – are upregulated in mammalian embryonic stem cells[6]. This upregulation in stem cell stages seems to have evolved convergently[6].

This study builds on the hypothesis that TEs act like EVEs in regulating stem cell identity, and are involved in the regenerative function of animals. It aims to identify putative regeneration-related TEs, annotate them, and infer their role in regeneration with bioinformatic approaches. It then uses wet-lab experiments to attempt to establish causal relationships between TEs and regenerative ability. It also compares the TE profiles surrounding regeneration-related genes in closely related vertebrate species with different regenerative ability. Through this comparative study, we can elucidate the role that specific TEs play in the regeneration of certain model vertebrates and gain insights on whether the infiltration and amplification of TEs in the genome in the vertebrates' evolutionary history have contributed to the change of their regenerative function.

This proposed study is significant in multiple ways. Because of the sequence similarity of TEs to each other, knowing the extent to which TEs regulate vertebrate regeneration would contribute to the future development of approaches to manipulate regeneration in both human and non-human vertebrates. Besides, given that TEs and EVEs possess radically different evolutionary dynamics from the host genome[7], studying these mobile genetic elements can shed light on the evolutionary process of the divergent regenerative ability in the vertebrates and the convergent evolution of the involvement of these elements in stem cell identity.

BACKGROUND

Since the discovery of the regenerative ability of *Hydra* in 1740, regenerative ability has been found throughout much of the animal kingdom[8]. However, this ability differs significantly among vertebrate lineages[4]. Given that regeneration ability is widespread among other metazoans, such as the annelids, the flatworms and the tunicates[8], it has long been debated if regeneration is an ancestral trait lost in evolution or an adaptive trait that has independently evolved for multiple times in the vertebrates[9]. While some teleost, amphibian and lizard species possess impressive abilities to regenerate whole appendages, birds and mammals have very limited ability to regenerate[4]. Despite this marked difference in regenerative ability, the genes that underlie regeneration in the vertebrates – mainly the same ones involved in embryonic development – are evolutionarily conserved, suggesting that these genes' regulation might be crucial to the difference of regenerative ability among the vertebrates[10,11]. Besides, the facts that there are few regeneration-specific genes[12], and that the same developmental genes are reactivated in some animals but not others following an injury[11], have further led to the speculation that regulatory elements play a significant role in the variation of tissue regeneration capacity across the animal kingdom[13,14]. Recently, regeneration-responsive enhancers were indeed revealed, and it was suggested that they play a role in shaping the regeneration abilities of different vertebrate species[15].

Apart from host mechanisms of gene regulation, TEs consist of an alternative group of elements that regulate gene expression. There is ample evidence supporting the involvement of TEs in host gene expression due to exaptation, such as contributing towards coding RNAs and providing *cis*-regulatory elements[1]. While

they interfere with host function in the process of replicating themselves, resulting in evolutionary arms races[16] and host defence mechanisms such as PIWI-interacting RNAs[17], they are genetically predisposed to recruit host transcription machinery due to the RNA polymerase promoter sequences within them[1]. Their sheer number and capacity to integrate into new genome loci also underlie their usefulness in rewiring gene regulatory pathways. They were found to regulate gene expression from animals to plants, controlling developmental switches and altering anthocyanin production[18]. Thus, it is likely that TEs constitute some of the regeneration-responsive enhancers in the genome, and that the regulatory control from TEs accounts for the highly variable regeneration capacity among vertebrate taxa to some degree[14,15].

In addition, TEs and EVEs are intimately linked with stem cell identity, a critical feature in more advanced animal regeneration abilities. Evidence on this association has been found in mammalian embryos: studies on human and mouse embryos suggest that several families of TEs and EVEs are active during early embryonic development[4] and in somatic stem cells[3]. Their activity is associated with cell pluripotency in mouse embryos[19], and this ability has convergently evolved in different lineages[7]. For instance, the activity of HERVH – a primate endogenous virus lineage – has been identified as a marker of stem cell function in humans[4]. The activity of a class of TEs, the LINE1 retrotransposons, has been found in mouse embryonic stem cells as well[5].

Such an association between these elements and stem cell identity might be due to epigenomic reprogramming, in which the resetting of the epigenomic status of the host genome – such as histone depletion and DNA hypomethylation – in early embryonic development offers an opportunity for TE propagation[3,20]. TEs seem to be making use of this period: some have evolved strong promoter function, possibly to maximally propagate themselves in this brief window – they would have a selective advantage if they can control and extend cell pluripotency[6]. These activities could have been exapted by the host for the early expression of host genes[20]. For example, epigenomic reprogramming directly leads to the transcription of full-length LINE1, a family of retrotransposon[1,3]. The methylation level of the DNA that codes for some regeneration-related genes in the somatic cells in some remarkably regenerative animals – such as the salamanders and the zebrafish – remains low after differentiation, providing a potential window similar to that of the embryonic stem cells in which TEs surrounding these genes might activate in a stem cell-like manner when tissue injury is perceived[21]. Given the heavy involvement of stem cells in vertebrate regeneration, it is reasonable to speculate that these mobile elements play a substantial role in regulating regeneration-related genes in animals.

Nevertheless, studies that associate TEs with the regeneration of evolutionarily distinct vertebrate lineages have been scarce. The presence of TE-derived RNA transcripts in regeneration transcriptomes – a hallmark of the activity of some TEs – has only been mentioned in a small number of studies, which found differential expression of TEs in the regenerating structures of these animals (Table 1). To our knowledge, there have been no studies so far that attempt to map and annotate putative regeneration-related TEs on the genome or to establish causal relationships between TEs and regeneration. Moreover, TE profiles are poorly conserved across species. The distinctive TE profiles among animal lineages are partly due to the occurrence of different virus endogenisation events in different lineages[2]. Thus, how these endogenisation events might relate to the evolution of regeneration in different vertebrate lineages remains an open question.

Species	Class of upregulated TE	Ref(s)
Axolotl (*Ambystoma mexicanum*)	LINE1 (Retrotransposon, germ cell marker)	22, 23
Iberian ribbed newt (*Pleurodeles waltl*)	Myb/Harbinger (DNA), Gypsy (LTR)	24
Sea cucumber (*Holothuria glaberrima*)	BEL (LTR), Gypsy (LTR)	25

Table 1. The TEs that are differentially expressed during the regeneration process of some animals. LTR = long terminal repeats, a subclass of TEs. DNA = DNA transposons.

Besides, despite the significant possibility of the relationship between TEs and regeneration discussed above, research and annotation on animals' TE profiles remain poor[2]. This disparity might have been partly due to the large number of genomes sequenced recently and the significant amount of manual labour required for annotating them[2]. However, the wide availability of resources in genomics and the advances in experimental techniques in high-resolution gene editing and gene expression control have provided the tools that allow us to explore this field more thoroughly[4]. In fact, the recent surge in the number of sequenced genomes serves as the basis of this study, without which it would be impossible to conduct.

These new genomes of regenerating animals would enable us to annotate previously undescribed putative regeneration-related TEs in the genome, providing general knowledge for future research in the field. It can also tell us more about the extent to which the involvement of TEs in embryonic development can be applied to other aspects in which stem cells play a significant role, such as regenerative medicine and the restoration of injured tissues in humans. Besides, it can potentially reveal how regenerative function has evolved or lost in different vertebrate lineages and how different TE subtypes are associated with regeneration, through which insights on manipulating vertebrate regenerative function can be made.

AIMS & OBJECTIVES

This study aims to elucidate the role that TEs play in vertebrate regenerative ability and its evolution. To investigate this aim, three main objectives have been set.

Objective 1: identify and annotate putative regeneration-related TEs in vertebrate genomes and transcriptomes. We would use currently available genomic and transcriptomic data in several regenerating and non-regenerating species to identify such TEs, annotate them, and compare their prevalence among species. On completion of this objective, we can make a list of candidate TEs involved, some of which would be experimentally tested in the following stage. We can also speculate about the mode of involvement of TEs in regeneration – i.e. whether their role involves single elements with great significance or multiple elements each with a smaller role – by mapping back their expression in the regeneration transcriptome to distinct genomic loci, for example. This hypothesis would also be tested at the following stage of the research.

Objective 2: establish causal relationships between TEs and their function in vertebrate regeneration. TEs have been found to wire new genes into the regulatory network of human embryonic stem cells[20]. Thus, we would like to investigate if their association with stem cell identity has been exapted to control cell pluripotency in vertebrate regeneration, and if they are causal agents of vertebrate regenerative ability, in wet-lab experiments with knockdown and knockout mutants of selected model species. We would also explore if the upregulation or silencing of specific families of TEs consists of a switch between normal and regenerative development. These experiments can elucidate the causal relationship between TEs and regenerative function and provide new approaches to stimulate or reactivate regeneration in different vertebrate lineages.

Objective 3: investigate the involvement of TEs in the evolution of regenerative function in the vertebrates. This objective builds upon the establishment of causal relationships between TEs and vertebrate regeneration in the previous objective. Much is still unknown about why regenerative ability differs dramatically among vertebrate lineages[8]. With the radically differing TE profiles revealed by the genomes of various species[23,24,26-32], we can speculate that the endogenisation events of viruses into and the spread of TEs in the genome in evolutionary history might have played a role in the change of regeneration genetic systems. In this objective, we would compare the presence and activity of regeneration-related TEs that we have found in Objectives 1 and 2 in groups of closely related species with different regenerative ability. Combining this with the genomes of more distant species, we can study the evolutionary history of these TEs, such as their possible origin and divergence. This information would also help us in future studies that attempt to induce regeneration in non-regenerating species by controlling the expression of specific TEs.

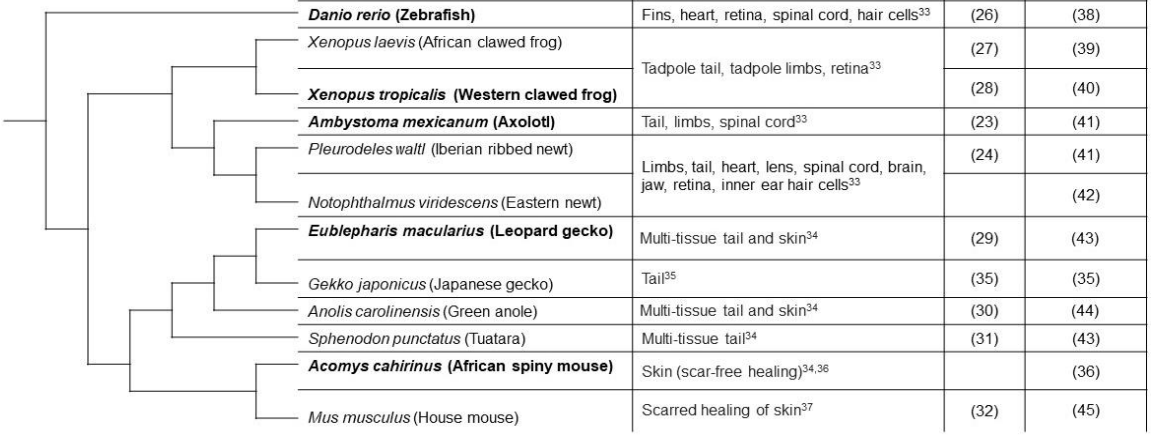
EXPERIMENTAL APPROACH

Identify and annotate putative regeneration-related TEs in vertebrate genomes & transcriptomes (Objective 1)

We would analyse the genomes and the transcriptomes of various model vertebrate species with significant regenerative ability and phylogenetically closely related species with differing regenerative abilities to facilitate inter-species comparison (Figure 1 shows the species proposed for research and their phylogeny).

The starting point of identifying these TEs is to locate and extract specific genomic regions containing regeneration-related genes. We would utilise the BLAST (Basic Local Alignment Search Tool) command line applications to isolate sequences upstream and downstream of the regeneration-related genes from genomic databases[47]. We would then download TE reference sequences from publicly available online sequence databases, such as Dfam[48], and search for putative TEs located close to regeneration-related genes using the HMMER software[49]. An example pipeline with these features has been developed (Figure 2). It is available

Figure 1. Phylogeny of proposed species and their regenerative ability. **Bold** – model species for experiments. Genomic and transcriptomic data available if cited. Phylogeny from (46).



Species	Regenerative parts	Data availability	
		Genome	Transcriptome
***Danio rerio* (Zebrafish)**	Fins, heart, retina, spinal cord, hair cells[33]	(26)	(38)
Xenopus laevis (African clawed frog)	Tadpole tail, tadpole limbs, retina[33]	(27)	(39)
***Xenopus tropicalis* (Western clawed frog)**		(28)	(40)
***Ambystoma mexicanum* (Axolotl)**	Tail, limbs, spinal cord[33]	(23)	(41)
Pleurodeles waltl (Iberian ribbed newt)	Limbs, tail, heart, lens, spinal cord, brain, jaw, retina, inner ear hair cells[33]	(24)	(41)
Notophthalmus viridescens (Eastern newt)			(42)
***Eublepharis macularius* (Leopard gecko)**	Multi-tissue tail and skin[34]	(29)	(43)
Gekko japonicus (Japanese gecko)	Tail[35]	(35)	(35)
Anolis carolinensis (Green anole)	Multi-tissue tail and skin[34]	(30)	(44)
Sphenodon punctatus (Tuatara)	Multi-tissue tail[34]	(31)	(43)
***Acomys cahirinus* (African spiny mouse)**	Skin (scar-free healing)[34,36]		(36)
Mus musculus (House mouse)	Scarred healing of skin[37]	(32)	(45)

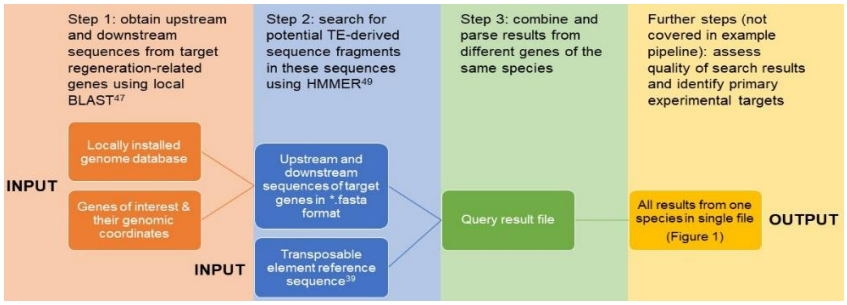


Figure 2. Flowchart of example pipeline to isolate putative regeneration-related TEs from genomic data.

from our GitHub repository, https://github.com/Anonymous-ox/te_regen. A list of genes and query results generated on 8 regeneration-related genes in *D. rerio* and a TE reference sequence from the Dfam database (hAT-N61_DR) is available on the same repository. The pipeline is readily applicable to other species.

The next step would be to select experimental targets from the list, in which different criteria are required for search results of genomic or transcriptomic origin as they can have distinctive mechanisms of action. For example, while genome-derived hits might act as promoters or enhancers of gene expression, transcriptome-derived hits might regulate gene expression by RNA-RNA or RNA-protein binding. We thus propose two separate sets of criteria to generate scores for our hits, thereby identifying experimental targets:

A. For transcriptome-derived results, e.g. those differentially expressed in regeneration transcriptome[22-25] –
   1) Relative activity in regenerating species compared to that in non-regenerating species;
   2) Significant expression in a high number of regenerating vertebrate species;
   3) Proximal location to regeneration-related genes and upregulation in the regeneration blastema.
B. For genome-derived results, e.g. potential promoters of regeneration-related genes –
   1) Prevalence in orthologous genomic regions of different regenerating vertebrates;
   2) Proximal location to regeneration-related genes;
   3) High affinity and conservation of any promoters they might have to different transcription factors.

We would generate a score for each hit and produce an ordered ranking for each of the three criteria (in either set A or B). We would then sum up the ranks of a hit in all three criteria and take the top ones for wet-lab experiments involving individual TEs. Upon completing this objective, we would have constructed a list of putative TEs involved in vertebrate regeneration, with its elements scored by a series of criteria. We would also have obtained specific TE families that are highly prevalent among the results. These results would help us identify experimental targets for the wet-lab experiments in the next objective.

Establish causal relationships between TEs and regenerative function (Objective 2)

This objective's accomplishment requires wet-lab experiments involving the manipulation of putative regeneration-related individual or family of TEs identified in the previous objective. Genetic engineering techniques, including CRISPR-Cas and RNA interference (RNAi), are the main methods in this objective. CRISPR-Cas and RNAi systems are fit for different purposes. With CRISPR-based gene-editing systems, we would focus on individual TEs of either genomic or transcriptomic origin. In contrast, RNAi allows us to potentially block the transcription of a whole group of TEs simultaneously – with the group size depending on the probe design – because of their sequence similarity. We mainly select well-studied model species (Figure 1) for this objective due to the ease to rear them and to generate transgenic lines. We would try to translate the significance of these results to the other proposed species in Objective 3.

Due to limited resources, we would provisionally identify 6 TEs from the top of our scoring system (3 genome-derived and 3 transcriptome-derived for CRISPR constructs) and 2 TE families with their members frequently identified among our pipeline's search results (for RNAi constructs) as our experimental targets for each species. RNAi constructs would be designed from consensus sequences for each TE family that we identify as an experimental target. We would then set up three treatment groups for each target, inserting the CRISPR or RNAi knockout/knockdown construct at different developmental stages – in the fertilised egg, the embryo limb, or before sexual maturity – and conduct amputation experiments at the designated developmental stages (Table 2). We would also set up three control groups to evaluate off-target effects – two with an empty CRISPR or RNAi construct inserted or transfected at the three above developmental stages respectively, and one with no construct (Table 3). We would closely monitor all groups to investigate these questions:

1) Do the constructs interfere with normal development and growth? (investigated by direct observation)
2) Is there a change in the expression of the experimental target, or the gene it is hypothesised to control, in the transcriptome when the animal regenerates after the amputation of the body part listed in Table 2?

	D. rerio	*X. tropicalis*	*A. mexicanum*	*E. macularius*	*A. cahirinus*
Developmental time	Adult	Tadpole	Adult	Adult	Adult
Body part	Fin	Tail	Limb	Tail	Skin

Table 2. Developmental time and body parts involved in amputation experiments for each species.

Target / Stage	CRISPR experiments							RNAi experiments			Wild-type control
	A	B	C	D	E	F	Control 1	G	H	Control 2	Control 3
Fertilised egg	1	4	7	10	13	16	19	22	25	28	31
Embryo limb	2	5	8	11	14	17	20	23	26	29	
Adolescent	3	6	9	12	15	18	21	24	27	30	
Sufficiency	32	33	34	35	36	37	38				

Table 3. CRISPR and RNAi experimental design. The number in the cells are group numbers.

(investigated by total RNA sequencing of cell samples from the animal's regenerating structure)

3) Do the constructs interfere with the animal's regeneration? (investigated by amputation experiments)

We would evaluate whether the observed changes in the animals' regeneration transcriptomes are significant in the treatment groups. We would also measure the regenerative ability (in measures of length of the regenerated appendages and time used) in each individual and perform statistical analyses, such as t-tests, on these results, to evaluate the extent to which the experimental target affects regeneration. We plan to take all experimental targets with statistically significant results found in questions 2 and 3 to the next objective.

The amputation experiments would show that the TEs are necessary for normal regenerative ability. Should the knockout of some TEs be found to affect regeneration, we would also reintroduce the experimental target in CRISPR-modified animal lines to test whether this is sufficient to restore the animal's regenerative ability. This experiment aims to establish causal relationships between individual TEs and regeneration (Table 3). However, it would not be done for RNAi targets since the nature of the RNAi construct's targets is unknown.

Investigate the involvement of TEs in the evolution of regenerative function in the vertebrates (Objective 3)

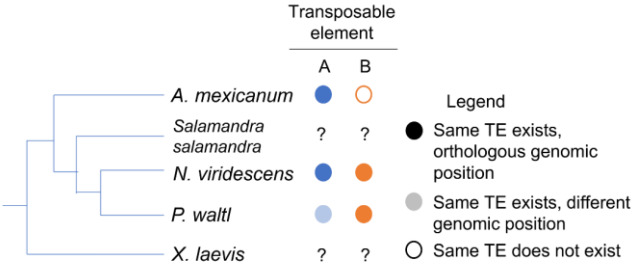
This objective builds on the regeneration-related TEs identified in the previous objective. Using the aforementioned bioinformatic approaches such as BLAST and HMMER, we can search for the same class of TEs in orthologous genomic positions in closely related species with different regeneration abilities (Figure 1). We can also utilise Biopython, a Python programming package, to organise and store these results[50].

Depending on the outcome of this search, several further steps would be taken (Figure 3). Firstly, the examined genomes may share the same TEs at orthologous genomic positions. If this is the case, we would utilise software such as Bowtie and TopHat to evaluate their activity by matching them with transcriptomic data[51]. If only genomic data is available, we can look for open reading frames or utilise Tajima's D or dN/dS statistics to elucidate their evolutionary history. Secondly, the same TEs might also be differentially located, expressed or regulated in a different species, in which case we would examine the genomes of more closely related species to investigate the TE's movement in its evolution. Thirdly, the same TEs might not exist in the other genomes that we analyse. In this case, we would be interested in investigating its evolutionary origin by looking at additional species' genomes. In future studies, we can also insert these TEs into orthologous genomic positions of closely related, non-regenerative species to further characterise their role in regeneration.

CONCLUSION

The proposed study investigates the relationship between transposable elements and vertebrate regeneration through a combination of computational and experimental approaches. The most crucial evidence – causal relationships between particular TEs and regenerative ability – would come from wet-lab experiments that alter TE activity and possibly the regenerative ability of treatment groups. The experimental targets for these experiments would come from bioinformatic searches within the genome of the model species. If causal relationships can be established, they would suggest that TEs do play a role in regulating regenerative activity in some vertebrates. After that, bioinformatic procedures can be resorted again, further extrapolating upon these elements and studying their functions and evolution among different vertebrate species. They can provide further evidence in the evolution of the relationship between TEs and regenerative ability in the vertebrates.

Figure 3. Some possible scenarios when we analyse regeneration-related TEs. TE A's causal relationship to regeneration was first identified in *A. mexicanum* (Scenarios 1 & 2), but is located at a different genomic position in *P. waltl.* In this case, we would be interested in investigating if A plays any function in *P. waltl* if at all and the evolutionary events that contributed to it, and whether restoring it to its orthologous position would affect the regeneration ability of *P. waltl*. We would also evaluate the potential activity of this TE in *N. viridescens* with bioinformatic approaches. The sequence of TE B exists in both *N. viridescens* and *P. waltl* but not in *A. mexicanum* (Scenario 3). In this case, we would be interested in tracing the evolutionary origin of B by looking at other genomes. Phylogenetic tree from (46).



Workplan for grant proposal *Comparative study of the potential role of transposable elements in vertebrate regeneration*

Year	Year 1				Year 2				Year 3			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4
Objective 1: characterise and annotate putative regeneration-related TEs in vertebrate genomes and transcriptomes.												
1.1 Identify the species and regeneration-related genes to investigate												
1.2 Search for putative TEs in genomes and transcriptomes and annotate them												
1.3 Develop a scoring system for putative TEs and utilise it to score search results												
1.4 Compare TE prevalence among species, analyse data and select experimental targets for the next objective												
Objective 2: establish causal relationships between TEs and their function in vertebrate regeneration.												
2.1 Construct and obtain transgenic lines of experimental species												
2.2 Conduct CRISPR knockout experiments for individual TEs												
- 2.2.1 Necessity experiments												
- 2.2.2 Sufficiency experiments on knockdown animal lines												
2.3 Conduct RNAi knockdown experiments for families of TEs												
2.4 Data analysis and target selection for Objective 3												
Objective 3: investigate the involvement of TEs in the evolution of regenerative function in the vertebrates.												
3.1 Analyse TE search results across species differing in regeneration abilities												
3.2 Conduct bioinformatic studies to investigate the evolution of the relationship between TEs and regeneration in multiple species												
3.3 Analyse data, make conclusions and point future directions such as wet-lab experiments involving trans-species TE transformation												
Milestones												
Generate a list of putative regeneration-related TEs for wet-lab experiments												
Discovery of association between TEs and regenerative ability												
Establishment of causation between TEs and regenerative ability												
Tracing of the evolutionary history of regeneration-causing TEs												

Justification of resources for project proposal *Comparative study of the potential role of transposable elements in vertebrate regeneration*

Only directly incurred costs (DIC) are included. Staff costs are not considered, although it is estimated that 2-3 lab technicians/postdocs would be optimal. The bulk of the cost would come from wet-lab experiments and sequencing. For each species, we propose to conduct the following: experimental groups of CRISPR (6 knockout targets and 6 for sufficiency experiments) and RNAi (2 knockdown targets) constructs inserted or transfected at three developmental stages (fertilised egg/embryo/adolescent); 7 control groups per species as outlined in Objective 2 of the Experimental Approach section in the main text. Thus, there would be 38 groups and 76 samples per species for transcriptome sequencing, if we allow 2 samples per group to be sequenced.

Computer hardware and server – A PC and appropriate additional local storage space (approximately 20TB) are necessary for storing a command line BLAST genomic database for convenient searches and regular updates. A multicore rack server is also requested for continuous access to fast and continuous access to computing. The data would be backed up regularly on the department's cloud server, and we assume that the department would be providing a server room for server storage. These pieces of equipment are critical for computational analysis. They provide enough RAM for conducting computational analysis, scoring specific TEs and storing computational and experimental results.

Item	Quantity required	Cost (£)	Sub-total
Personal computer (Intel i9 CPU – e.g. Dell Precision 5820 Tower)	1	2000	2,000
Monitors (e.g. Lenovo L28u-30)	2	254	508
Local storage space for genomic databases (2TB hard disk drives, each)	10	55	550
Multicore rack server (e.g. Dell PowerEdge R7525 Performance)	1	3775	3,775
Miscellaneous (adaptors, cables, etc.)			100
		Total	**6,933**

Transgenic constructs and sequencing – CRISPR knockout and RNAi knockdown constructs need to be designed and created to conduct wet-lab experiments on candidate individual TEs or TE families in 5 species to investigate their relationship with regeneration in these species. For CRISPR lines, we plan to utilise services from Genome Engineering Oxford (GEO) to create custom-made CRISPR constructs. We plan to use PAGE-purified custom Gene Link™ RNAi transfection constructs for our RNAi experiments. We are also proposing to sequence the regeneration transcriptome of the experimental animal (2 individuals per group) and examine candidate TE expression. RNA extraction, library construction and sequencing is essential for the study.

Item	Quantity required	Cost (£)	Sub-total
CRISPR constructs (GEO) (12 experimental & 2 empty control per species, needed for both knockout and sufficiency experiments)	70	95	6,650
RNAi constructs (e.g. PAGE-purified custom Gene Link™ RNAi construct, 10 nmol) (2 experimental & 1 empty control per species)	15	400	6,000
Total RNA extraction reagent (e.g. Invitrogen™ TRIzol™ Plus RNA Purification Kit) for 50 preps	7	283	1,981
High-throughput total RNA sequencing (each sample)	380	245	93,100
		Total	**107,731**

Experimental animal supplies and transgenic line generation – We allot 200 embryos per group for fish and amphibian species and 10 per group for geckos. To save time and efforts, we plan to utilise commercially available custom transgenic lines if available and construct our transgenic line if this is not possible (this is likely to be the case for *A. mexicanum*, *E. macularius* and *A. cahirinus*). The provisional cost to generate transgenic lines in our laboratory is included in the next section. These transgenic lines, containing knockdowns of candidate TEs in this study, are essential for elucidating the causal relationship between TEs and regeneration in different vertebrates.

Item	Quantity required	Cost (£)	Sub-total
D. rerio WT embryos x100 (Zebrafish International Resource Center)	76	40	3,040
D. rerio custom transgenic line (InVivo Biosystems)	10	751	7,510

X. tropicalis WT embryos x100 (Marine Biological Laboratory, Chicago)	76	80	6,080
X. tropicalis custom transgenic line	10	940	9,400
A. mexicanum WT embryos x100 (Ambystoma Genetic Stock Center, University of Kentucky)	76	46	3,496
E. macularius eggs for lentivirus transgenic construct[52]	380	10	3,800
Lentivirus transgenic design (VectorBuilder)	10	1127	11,270
Other costs (empty CRISPR constructs & viral vectors for control groups, pairs of adult *E. macularius* and *A. cahirinus* for breeding, etc.)			5,000
		Total	**49,596**

Equipment and consumable laboratory supplies – We expect the availability of general equipment in a developmental biology laboratory. Besides, we request a camera to record phenotypic results in detail. As we plan to generate at least some CRISPR and RNAi transgenic lines, there would be considerable costs and efforts associated with this activity. We expect having to develop transgenic lines for 3 species ourselves, and the cost of generating transgenic lines for one species (such as *A. cahirinus*) has been estimated from that of generating a transgenic mice line[53]. An additional CO_2 incubator is required to rear embryos and an inverted microscope for embryo microinjection. Besides, funds are also requested for other consumables such as disposable plasticware & metal, personal protective equipment, reagents, solvents, glass capillary tools, etc.

Item	Quantity required	Cost (£)	Sub-total
Microscopes – brightfield microscopes, stereo microscopes	2 each	(Provided)	
Camera	1	500	500
Inverted Hoffman Modulation Contrast Microscope (Olympus IX71)	1	12700	12,700
CO_2 incubator (Thermo Scientific™ Midi CO_2 incubator, 40L)	1	4254	4,254
Consumables: transgenic construct reagents (per species)	3	4621	13,863
Other consumables (per year)	2	1000	2,000
		Total	**33,317**

Growth facility and maintenance – We assume that we have access to one environmentally-controlled room for each model species we use (5 rooms required in total). It is expected that approx. 200 embryos would fit one tank for amphibian and fish (as not all would develop), but this would be adjusted according to the actual outcome of the animals' development[54]. For *E. macularius* we allot each pair a tank with environmental setups. Since *A. cahirinus* live in groups of 10-20, one environmentally-controlled room with 20 cages would fit the number required for experimental line generation[55]. The per-unit cost to rear each species would differ among species but, due to space limits, they are not listed separately but provided as the average for all species; the total cost would be the per-unit cost multiplied by the number of species used. Necessary funds to rear these experimental animals (food, environmental control, enrichments, etc.) are outlined in this section.

Item	Quantity required	Cost (£)	Sub-total
Rearing equipment (e.g. water & air recirculating systems, light regimes, embryo incubators) (per species)	5	500	2,500
Additional tanks (60 * 60 * 30 cm) for amphibians, fish, and geckos	124	60	7,440
Wired cages (60 * 45 * 40 cm) for mice	20	60	1,200
UV steriliser and environmental condition monitors (per species)	5	200	1,000
Food items (per species per year)	10	300	3,000
Environmental enrichments (e.g. plastic plants, rocks, temperature regulators), management equipment (e.g. siphons, nets) (per species)	5	600	3,000
Surgical and dissection instruments			1,000
Miscellaneous costs: shipping charges, international document fees, health certificates, etc.			3,000
		Total	**22,140**

Travel and impact-related activities – funds are requested for the attendance of two academic conferences (at £500 each) and the development of a video and online resources to highlight the results and significance of this project (£2,000). **Section total = £3,000.**

Total estimated cost = £222,717.

Data management plan for project proposal *Comparative study of the potential role of transposable elements in vertebrate regeneration*

Data areas and data types

In silico analyses: this study would utilise publicly available genomic and transcriptomic data, such as reference TE sequences (in statistical models such as *.hmm and *.sto), genome assemblies for database construction (downloaded from the NCBI in *.fna format) and transcriptome assemblies. Python scripts would be written for the pipelines to select experimental targets. The analyses would also generate annotations of genomic sequences, which would be kept as GenBank format files. *.fasta formats for sequences and *.csv spreadsheets are used to store intermediate and final results of computational analyses. The sequencing of animal samples would produce total transcriptome datasets. Various softwares would also be used to investigate the evolution of regeneration-related transposable element sequences, and data would be transformed and curated to suit the data types they are compatible with.

In vivo experiments: wet-lab experiments would generate photographs of experimental results in macroscopic and microscopic detail, time-lapse videos of embryonic development and organ regeneration, body part measurements in amputation experiments, as well as laboratory notes on experiments and routine animal husbandry. They would be annotated if necessary, stored in a dedicated computer, backed up regularly, and uploaded to a repository. These experiments would also generate CRISPR or RNAi construct designs and transgenic animal lines with these constructs.

Standards and metadata

The data generated in the scope of the project would be stored in appropriate formats for more convenient future use and assuring the reproducibility of the study. For instance, sequence metadata would be stored and managed in GenBank format files. They would also be backed up on storage facility owned by the department regularly. The data to be shared publicly would contain metadata for explanations and clarifications if appropriate, and we would ensure that they comply with BBSRC's Data Sharing Policy.

Relationship to other data available in public repositories

The primary data source of this study is publicly available genomic data and transposable element reference sequences. Genome and transcriptome annotations that result from this study would be made publicly available for the interest of other researchers.

Secondary use

Completed datasets would be made available through publicly accessible file repositories in their appropriate form. Codes of the analysis pipelines will include basic annotations. These datasets can assist with the future development of similar pipelines to analyse genomic sequences. The use of quantitative results regarding the interactions between genes and TEs would be used in further experimental research proposed by this project, such as using the results to construct transgenic lines to restore regeneration in non-regenerative animals as outlined in Objective 3. This project would also generate transgenic animal lines, which would be reared in our laboratory and can be distributed upon request, mutual agreement, and standard procedures.

Methods for data sharing

The principle is to make all data as accessible as possible. The annotations generated from this project would be uploaded onto GenBank. Code used for the pipeline would be made available through GitHub, a major sharing platform for coders, while other files would be uploaded onto public repositories such as Figshare and Dryad Digital Repository. Animal lines, transgenic constructs, and other materials generated by wet-lab experiments will be distributed upon request, mutual agreement, and standard procedures.

Proprietary data

No data collected as part of the project will be subject to proprietary data restrictions.

Timeframes

The types of data mentioned above will be publicly released in their entirety when the project is concluded through the methods outlined above in the Methods for Data Sharing section.

References

1. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, et al. *Genome Biol*. 2018;19(1):199. **2.** Sotero-Caio CG, Platt RN 2nd, Suh A, Ray DA. *Genome Biol Evol*. 2017;9(1):161-177. **3.** Schumann GG, Fuchs NV, Tristán-Ramos P, Sebe A, Ivics Z, Heras SR. *Mob DNA*. 2019;10:9. **4.** Hutchins E, Kusumi K. (2016). Genetics and regeneration in vertebrates. In Steinhoff G (2016), *Regenerative Medicine – from Protocol to Patient: 1. Biology of Tissue Regeneration*. Switzerland: Springer. **5.** Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, et al. *Cell*. 2018;174(2):391-405. **6.** Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. *Bioessays*. 2016;38(1):109-117. **7.** Feschotte C, Gilbert C. *Nat Rev Genet*. 2012;13(4):283-296. **8.** Tanaka EM, Ferretti P. *Nat Rev Neurosci*. 2009;10(10):713-723. **9.** Lai AG, Aboobaker AA. *Dev Biol*. 2018;433(2):118-131. **10.** Sebé-Pedrós A, Ballaré C, Parra-Acero H, Chiva C, Tena JJ, et al. *Cell*. 2016;165(5):1224-1237. **11.** Rodriguez AM, Kang J. *Semin Cell Dev Biol*. 2020;97:47-54. **12.** Goldman JA, Poss KD. *Nat Rev Genet*. 2020;21(9):511-525. **13.** Suzuki N, Ochi H. *Dev Growth Differ*. 2020;62(5):343-354. **14.** King BL, Rosenstein MC, Smith AM, Dykeman CA, Smith GA, Yin VP. *NPJ Regen Med*. 2018;3:10. **15.** Wang W, Hu CK, Zeng A, Alegre D, Hu D, et al. *Science*. 2020;369(6508):eaaz3090. **16.** Koonin EV, Makarova KS, Wolf YI, Krupovic M. *Nat Rev Genet*. 2020;21(2):119-131. **17.** Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. *Nat Rev Genet*. 2019;20(2):89-108. **18.** Chuong EB, Elde NC, Feschotte C. *Nat Rev Genet*. 2017;18(2):71-86. **19.** Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, et al. *Nature*. 2012;487(7405):57-63. **20.** Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, et al. *Nat Genet*. 2010;42(7):631-634. **21.** Katsuyama T, Paro R. *FEBS Lett*. 2011;585(11):1617-1624. **22.** Zhu W, Kuo D, Nathanson J, Satoh A, Pao GM, et al. *Dev Growth Differ*. 2012;54(7):673-685. **23.** Nowoshilow S, Schloissnig S, Fei JF, Dahl A, Pang AWC, et al. *Nature*. 2018;554(7690):50-55. **24.** Elewa A, Wang H, Talavera-López C, Joven A, Brito G, et al. *Nat Commun*. 2017;8(1):2286. **25.** Mashanov VS, Zueva OR, García-Arrarás JE. *Mob Genet Elements*. 2012;2(5):244-247. **26.** Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, et al. *Nature*. 2013;496(7446):498-503. **27.** Session AM, Uno Y, Kwon T, Chapman JA, et al. *Nature*. 2016;538(7625):336-343. **28.** Hellsten U, Harland RM, Gilchrist MJ, Hendirx D, Jurka J, et al. *Science*. 2010;328(5978):633-636. **29.** Xiong Z, Li F, Li Q, Zhou L, Gamble T, et al. *Gigascience*. 2016;5(1):47. **30.** Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, et al. *Nature*. 2011;477(7366):587-591. **31.** Gemmell NJ, Rutherford K, Prost S, Tollis M, Winter D, et al. *Nature.* 2020;584(7821):403–409. **32.** Mouse Genome Sequencing Consortium. *Nature*. 2002;420(6915):520-562. **33.** Poss KD. *Nat Rev Genet*. 2010;11(10):710-722. **34.** Jacyniak K, McDonald RP, Vickaryous MK. *J Exp Biol*. 2017;220(Pt 16):2858-2869. **35.** Liu Y, Zhou Q, Wang Y, Luo L, Yang J, et al. *Nat Commun*. 2015;6:10033. **36.** Mamrot J, Legaie R, Ellery SJ, Wilson T, Seemann T, et al. *Sci Rep*. 2017;7(1):8996. **37.** Yoon JH, Cho K, Garrett TJ, Finch P, Maden M. *Sci Rep*. 2020;10(1):166. **38.** Cao J, Navis A, Cox BD, Dickson AL, Gemberling M, et al. *Development*. 2016;143(2):232-243. **39.** Sosa EA, Moriyama Y, Ding Y, Tejeda-Muñoz N, Colozza G, De Robertis EM. *Int J Dev Biol*. 2019;63(6-7):301-309. **40.** Tan MH, Au KF, Yablonovitch AL, Willis AE, Chuang J, et al. *Genome Res*. 2013;23(1):201-216. **41.** Joven A, Elewa A, Simon A. *Development*. 2019;146(14):dev167700. **42.** Looso M, Preussner J, Sousounis K, Bruckskotten M, Michel CS, et al. *Genome Biol*. 2013;14(2):R16. **43.** Tzika AC, Ullate-Agote A, Grbic D, Milinkovitch MC. *Genome Biol Evol*. 2015;7(6):1827-1841. **44.** Hutchins ED, Markov GJ, Eckalbar WL, George RM, King JM, et al. *PLoS One*. 2014;9(8):e105004. **45.** He P, Williams BA, Trout D, Marinov GK, Amrhein H, et al. *Nature*. 2020;583(7818):760-767. **46.** OneZoom Team. OneZoom Tree of Life Explorer. https://www.onezoom.org/. Published 2012. **47.** BLAST® Help. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/books/NBK1762/. Published 2008. Accessed January 3, 2021. **48.** Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, et al. *Nucleic Acids Res*. 2016;44(D1):D81-D89. **49.** Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, et al. *Nucleic Acids Res*. 2015;43(W1):W30-W38. **50.** Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, et al. *Bioinformatics*. 2009;25(11):1422-1423. **51.** Langmead B, Trapnell C, Pop M, Salzberg SL. *Genome Biol*. 2009;10(3):R25. **52.** Kaitlyn H, Dee H, Bob C, Hickok TW, Petitte JN, Mozdziak PE. *Afr J Microbiol Res.* 2014;8(10):1070-1079. **53.** Kumar TR, Larson M, Wang H, McDermott J, Bronshteyn I. *Methods Mol Biol*. 2009;590:335-362. **54.** McNamara S, Wlizla M, Horb ME. *Methods Mol Biol*. 2018;1865:1-17. **55.** Haughton CL, Gawriluk TR, Seifert AW. *J Am Assoc Lab Anim Sci*. 2016;55(1):9-17.