RESEARCH ARTICLE | JANUARY 10 2019

Conservation analysis showing high variability in the rate of evolution of regeneration-related genes FREE

Shuyan Zhou

AIP Conf. Proc. 2058, 020045 (2019)
https://doi.org/10.1063/1.5085558

Conservation Analysis Showing High Variability in the Rate of Evolution of Regeneration-related Genes

Shuyan Zhou[1, a)]

[1]*Lexington Christian Academy, Massachusetts 02420, the United States.*

[a)]Corresponding author email: 2310645312@qq.com

Abstract. Schmidtea mediterranea has a high regeneration ability. Previous studies have discovered genes contributing to this species' regeneration ability, including genes with orthologs in human. However, no previous comparative studies have investigated these orthologs across a diversity of species. Understanding how the regeneration-related genes evolved in a diverse range of extant species is helpful to both the biological and medical field. Here I reported conservation measurements, pairwise dN/dS, of regeneration-related genes generated from 12 species against human. I then compared them to the pairwise dN/dS of housekeeping under the same criteria and found that the evolution process of genes related to regeneration varies across species and was less conserved than the housekeeping genes, indicating a less evolutionary constraint over regeneration-related genes.

Key words: AIP Proceedings; Conservation Analysis Showing High Variability in the Rate of Evolution of Regeneration-related Genes; International Conference.

INTRODUCTION

DN is the rate of nonsynonymous substitutions at nonsynonymous sites, where the sequences are usually under more evolutionary constraints, and dS is the rate of synonymous substitutions at synonymous sites, which are assumed to undergo neutral evolution (Kryazhimskiy & Plotkin, 2008). By calculating dN/dS, we can get an accurate parameter to understand the evolution mode of a group of genes between a pair of species. The values of dN/dS were usually used to analyze the different genetic sequences of divergent species to understand certain patterned events in independent lineages (Goldman & Yang, 1994; KIMURA, 1977; Muse & Gaut, 1994). To understand to what degree the same genes related to regeneration evolve among different species, I compared the average dN/dS of a group of selected regeneration-related genes to the average dN/dS of the housekeeping genes.

Regeneration exist in many species, including S. med, human, cat, and mouse. Each specie has its regeneration capacity fitted for its survival in the natural world. Planarians live off their regeneration ability after serious damage to their bodies by growing back almost half of their body parts. However, the regeneration ability of human is relatively limited to regrowth of skin, hair, liver, and spleen. Because different species rarely share the same regeneration capacity, my hypothesis for this research was that the average dN/dS of the regeneration-related genes would be less conserved than the housekeeping genes.

Schmidtea mediterranea, a freshwater planarian that has outstanding regeneration ability controlled by pluripotent stem cells, is an ideal model for studying regeneration-related genes (Önal et al., 2012). To gain a further understanding of the evolution of regeneration-associated genes, it is important to know which genes are expressed in cells as the key factors for regeneration and to what degree the human share the similar sets of genes. Previous studies have verified several neoblast markers, including TGFBR2 and the targets of OCT4 (Önal et al., 2012; Reddien, Bermange, Murfitt, Jennings, & Sánchez Alvarado, 2005a). Besides those genes associated with the neoblast formation, specific genes like NANOG and SOX2, which are highly associated with the species' regeneration capacity and had only been studied in mammals, have been verified to also exist in planarians (Önal et al., 2012). Based on those evidence, I

International Conference on Frontiers of Biological Sciences and Engineering (FBSE 2018)
AIP Conf. Proc. 2058, 020045-1–020045-6; https://doi.org/10.1063/1.5085558
Published by AIP Publishing. 978-0-7354-1786-1/$30.00

focused on the most well-known and representative regeneration-related genes of Schmidtea mediterranea and translated those gene names into the same human genes for the later purpose of comparing the pairwise dN/dS of human and other species.

METHODS

Data Acquisition

The regeneration-related genes required for regeneration in this paper were acquired from several papers on planarians regeneration (Önal et al., 2012; Oviedo & Levin, 2007; Palmisano & Di Giovanni, 2018; Reddien, Bermange, Murfitt, Jennings, & Sánchez Alvarado, 2005b) [Figure 1]. Due to a variety of gene naming methods used in the papers, it was effective and significant to find an agreed S.med gene name. I used the data base of PlanMine (Brandl et al., 2016) (data accessed on Sep 5, 2018) to confirm the S.med gene names. From PlanMine, I downloaded the lists of all gene names of S.med under the QueryBuilder sub section by using the keywords Schmidtea mediterranea and bio entity. I sorted the gene names and clean the extra information with Unix bash commands to feed the Ensembl with machine readable information in the next step. I coded to strip away unnecessary information and finalize this list to only contain the gene names. The code can be found in GitHub (https://github.com/Emily-zsy/regeneration-related-genes-evolution-history.git).

With this sorted S. med gene list, I checked the profile of the genes in S.med and was able to find the genes with orthologs in humans in Uniprot (Bateman et al., 2017) (data accessed on Sep 13,2018). During this process, genes without human orthologs were tossed away, and repetitive genes were combined. The final list for genes needed for regeneration had 28 verified genes. To create a gene control group, I then selected a list for human housekeeping genes, which contained 11 genes, from literature (Eisenberg & Levanon, 2014).

To get the dN/dS of the two lists of genes between human and other species, I used the Ensembl Genes 93 database (data accessed on Sep 18, 2018). The lists of input genes were selected with two parameters: gene name and gene description. The twelve species were shown with four parameters: homology type, dN, dS, and ortholog confidence. Two types of outputs were generated: the dN and dS values of the housekeeping genes and the regeneration-related genes. Because Ensembl website only allows a submission of comparing human with 6 other species at one time, I put together the data of 12 species when uploading the data into Excel 2018. Based on the result from Ensembl, I calculated the dN/dS value of each gene in twelve species. Average dN/dS value were then calculated in two ways. First, the average was taken across the species, which means each specie had an average dN/dS score for all its genes of interest. Second, the average was calculated for each gene across different species, which means each gene of interest had an average dN/dS score for all 12 species.

One thing that should be taken carefully about the calculation of dN/dS is the treatment of 0/0 division. 0/0 appeared multiple times in my data set, such as SOX2 in Cat, EDF1 in Cow, and SMAD4 in Capuchin. Since 0 cannot be the denominator, 0/0 was artificially coded to be null (0/1). In this case, all 0/0 were removed from the lists while the average dN/dS were calculated.

Data Analysis

I selected the genes needed for regeneration for this research from five sources of literature. Before analyzing the average dN/dS values of those genes, it was clear to see that to what proportion each literature contributed my gene list of interest. I used the software BioVinci 2018 to generate the graph [Figure 1] (data accessed on Oct 26, 2018).

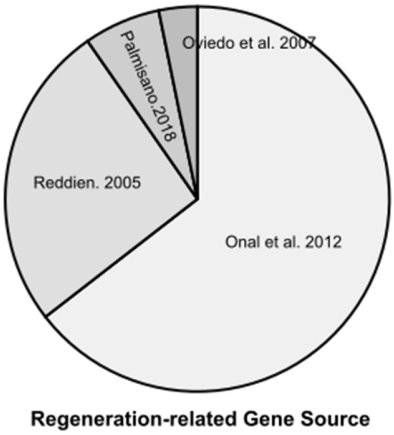

FIGURE 1. Selected S.med Genes from Different Sources of Literature. The genes selected for this research came from the five sources. The area on the pie chart shows the proportional contribution of the source to the genes in this research. Larger area demonstrates more genes of this research came from the source.

Both the average dN/dS of regeneration-related genes and housekeeping genes across species were analyzed under the Wilcoxon Signed Ranks test. The four assumptions of the test were met: the two samples were dependent observations of the cases; the paired observations were random and independent; the measurements of dN/dS were assumed continuous in theoretical nature; both the dependent measurements were of an ordinal scale. I used the software SPSS Statistics to run Wilcoxon Signed Rank test here since it was a non-parametric statistical hypothesis test for two paired samples, which fitted my research well (test run on 16 October 2018 20:38:02).

RESULT

From five sources of literature, I finalized two lists of human orthologs of S.med genes used for studying: one for the genes required for regeneration and the other for housekeeping genes [Table 1 and Table 2]. The list of genes needed for regeneration contains 28 genes: CSAG2, GNL3, DDX54, SMAD4, ACTL6A, WWP2, TGFBR1, TGFBR2, POU5F1, SOX2, NANOG, SMARCC2, CTR9, TDRD9, DAZL, TAF7, EIF2S2, EIF2B4, EDF1, TPR, NUP107, PCF11, NCBP1, SF3A1, SF3A3, KMT2A, PAF1, KMT5A [Table 1]. The list of housekeeping genes has 11 genes: C1orf43, CHMP2A, EMC7, GPI, PSMB2, PSMB4, RAB7A, REEP5, SNRPD3, VCP, VPS29 [Table 2].

TABLE 1. The Average dN/dS of Regeneration-Related Genes. This list of genes are the human orthologs of 28 regeneration-related genes found in S.med. The average dN/dS scores reported here is the average of pair-wise dN/dS between human and 12 other species of interest. The gene CSAG2 was not found in all the 12 species.

Regeneration genes	Average gene dN/dS	Regeneration genes	Average gene dN/dS
NUP107	0.1225647778	SMARCC2	0.0501936613
DDX54	0.09056241749	TGFBR1	0.07234203158
KMT5A	0.2136899548	NCBP1	0.01985100208
WWP2	0.04977052265	GNL3	0.4204590499
SOX2	0.01627232506	SF3A1	0.01832603274
ACTL6A	0.03636207997	CSAG2	N/A
TGFBR2	0.08339605836	TPR	0.05443683871
PAF1	0.02818337219	SMAD4	0.0239609844
TAF7	0.02828048612	DAZL	0.2442095425
SF3A3	0.03119039479	NANOG	0.4065479837
EIF2B4	0.2351504713	POU5F1	0.1105538475
KMT2A	0.1220915925	PCF11	0.09115838186
TDRD9	0.1510824292	CTR9	0.03152032227
EIF2S2	0.06524860531	EDF1	0.01808011651

TABLE 2. The Average dN/dS of Housekeeping Genes. This list of genes are the human orthologs of 11 regeneration-related genes found in S.med. The average dN/dS scores reported here is the average of pair-wise dN/dS between human and 12 other species of interest. The gene SNRPD3 was not found in all the 12 species.

Housekeeping genes	Average gene dN/dS	Housekeeping genes	Average gene dN/dS
VPS29	0.187323409	EMC7	0.08795421191
GPI	0.09968617144	VCP	0.02310851571
RAB7A	0.01038252934	SNRPD3	N/A
C1orf43	0.06772628261	CHMP2A	0.006142077804
PSMB2	0.04034285091	REEP5	0.07821926179
PSMB4	0.08433243258		

The average dN/dS, taken across the 12 species, of regeneration-related gene ranged from 0.016272325 (SOX2) to 0.406547984 (NANOG). The standard deviation of the average dN/dS of the regeneration-related genes was 0.03775982827 (variance=0.001425804631) [Table 3]. The average dN/dS, taken across the 12 species, of housekeeping genes ranged from 0.010382529 (RAB7A) to 0.187323409 (VPS29). The standard deviation of the average dN/dS of the housekeeping genes was 0.03598553556 (variance=0.001294958769) [Table 3]. In comparison, the average dN/dS of the regeneration-related genes were of a larger variance.

TABLE 3. The Pair-wise dN/dS scores of the Regeneration-related Genes and Housekeeping Genes Averaged across the Twelve Species. The variance of the average dN/dS of the regeneration-related genes was 0.001425804631, while the variance of the average dN/dS of the housekeeping genes was 0.001294958769. In comparison, the average dN/dS of the regeneration-related genes were of a larger variance.

Species	Average gene dN/dS for regeneration-related genes	Average gene dN/dS for housekeeping genes
Cat	0.08547434019	0.06892675827
Dog	0.1010128575	0.06087106835
Capuchin	0.1734794312	0.1041662185
Armadillo	0.08965360903	0.08935085844
Cow	0.1044571475	0.05655624885
Chinese hamster CriGri	0.07877800132	0.07118653879
Algerian mouse	0.07087458087	0.04765127009
Mouse	0.08874707414	0.04642638881
Squirrel	0.09398631856	0.03944710558
Angola colobus	0.1916780448	0.1697697662
Microbat	0.1003830758	0.07451620495
Damara mole rat	0.08072006515	0.04603005782
Variance	0.001425804631	0.001294958769

A Wilcoxon Signed Rank test on the two paired lists of average [Table 3] shows a great significance on the less conserveness on the regeneration-related genes to the housekeeping genes ($Z=-3.059$, $P<0.002$, two-way Wilcoxon Signed Rank Test; Table 4-6). This result showed that when compared with housekeeping genes, which were presumed to be more conserved, the evolution of regeneration-related genes was of greater changes [Table 4-6]. This result [Table 4-6], together with the previous evidence of the high variance in the average dN/dS of regeneration-related genes [Table 3], suggested that regeneration-related genes across species underwent significantly more changes during the evolution than the housekeeping genes.

TABLE 4. Wilcoxon Signed Ranks Test (Descriptive Statistics).

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
regeneration	12	.1049370455050	.03775982826715	.07087458087	.19167804480
housekeeping	12	.0729082070542	.03598553555794	.03944710558	.16976976620

TABLE 5. Wilcoxon Signed Ranks Test (Ranks). Housekeeping < regeneration; b. housekeeping > regeneration; c. housekeeping = regeneration)

		N	Mean Rank	Sum of Ranks
housekeeping - regeneration	Negative Ranks	12[a]	6.50	78.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	12		

TABLE 6. Wilcoxon Signed Ranks Test (Test Statistics). When Z=-3.059, P<0.002, the test shows that the evolution of regeneration-related genes was of greater changes when compared with housekeeping genes, which were presumed to be more conserved.

	housekeeping - regeneration
Z	-3.059[a]
Asymp. Sig. (2-tailed)	.002

(a. Based on positive ranks.)

DISCUSSION

In summary, I compared the conserveness of the representative regeneration-related genes, which I got from various sources of literature, across species. The housekeeping genes were used as a control group, representing highly conserved genes, and the genes needed for regeneration would be studied as the experimental group. The results answered the critical question raised on the topic of regeneration: to what extend the regeneration-related genes evolve in our world? In accordance with the data, those genes, which exist in multiple species, represent a less conserved trend of evolution. This conclusion from the study coincides with my hypothesis. Though most of the genes needed for regeneration are shared across species, their evolution was much less conserved to housekeeping genes, like GPI which codes for glycolytic enzyme to maintain the basic functions of human body. This is because of the phenotypic and mechanical difference between species.

As we notice this less conserved trend of the evolution of regeneration-related genes, more questions are going to emerge after this study. For the highly conserved genes like SNRPD3 and RAB7A, future research can investigate model organs to find out the reasons for their high conserveness. Besides this possible direction of research, another highly possible direction of future research is the possible clade-specific patterns of pairwise dN/dS against human. I found a high interspecies variability of average dN/dS of regeneration-related genes, but not in housekeeping genes. A diverse set of species from different lineages will possibly points out some pattern of the average dN/dS of regeneration-related genes, which would suggest different evolution patterns of genes required for regeneration in different clades, including the order primates.

REFERENCES

1. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., … Zhang, J. (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Research, 45(D1), D158–D169. Https: //doi. Org/ 10. 1093/ nar/ gkw1099.
2. Brandl, H., Moon, H., Vila-Farré, M., Liu, S.-Y., Henry, I., & Rink, J. C. (2016). PlanMine – a mineable resource of planarian biology and biodiversity. Nucleic Acids Research, 44(D1), D764–D773. Https: // doi. Org/ 10.1093/ nar/ gkv1148.
3. Eisenberg, E., & Levanon, E. Y. (2014). Corrigendum to: Human housekeeping genes, revisited. Trends in Genetics, 30(3), 119–120. https://doi.org/10.1016/j.tig.2014.02.001.

4. Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution, 11(5), 725–736. Https: // doi. Org/ 10.1093/ oxfordjournals. molbev. a040153.
5. KIMURA, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature, 267(5608), 275–276. https://doi.org/10.1038/267275a0.
6. Kryazhimskiy, S., & Plotkin, J. B. (2008). The Population Genetics of dN/dS. PLoS Genetics, 4(12), e1000304. https://doi.org/10.1371/journal.pgen.1000304.
7. Muse, S. V, & Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Molecular Biology and Evolution, 11(5), 715–724. https://doi.org/10.1093/oxfordjournals.molbev.a040152.
8. Önal, P., Grün, D., Adamidi, C., Rybak, A., Solana, J., Mastrobuoni, G., Rajewsky, N. (2012). Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. The EMBO Journal, 31(12), 2755–2769. https://doi.org/10.1038/emboj.2012.110.
9. Oviedo, N. J., & Levin, M. (2007). smedinx-11 is a planarian stem cell gap junction gene required for regeneration and homeostasis. Development, 134(17), 3121–3131. https://doi.org/10.1242/dev.006635.
10. Palmisano, I., & Di Giovanni, S. (2018). Advances and Limitations of Current Epigenetic Studies Investigating Mammalian Axonal Regeneration. Neurotherapeutics, 15(3), 529–540. https://doi.org/10.1007/s13311-018-0636-1.
11. Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R., & Sánchez Alvarado, A. (2005a). Identification of Genes Needed for Regeneration, Stem Cell Function, and Tissue Homeostasis by Systematic Gene Perturbation in Planaria. Developmental Cell, 8(5), 635–649. https://doi.org/10.1016/j.devcel.2005.02.014.
12. Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R., & Sánchez Alvarado, A. (2005b). Identification of Genes Needed for Regeneration, Stem Cell Function, and Tissue Homeostasis by Systematic Gene Perturbation in Planaria. Developmental Cell, 8(5), 635–649. https://doi.org/10.1016/j.devcel.2005.02.014.