Analysis of convergent and parallel amino acid substitutions in the HSP90AA1 gene among high-elevation anurans

Joyce Tao

Department of Integrative Biology, College of Biological Sciences, University of Guelph, Guelph, ON Canada. Faculty supervisor: Dr. Jinzhong Fu.
For correspondence, please email: joycetao3100@yahoo.com

Abstract

A significant amount of convergent and parallel amino acid substitutions in the HSP90AA1 gene has been detected among four species of high-elevation anurans: *Bufo tibetanus*, *Scutiger boulengeri*, *Rana kukunoris*, and *Nanorana parkeri*. As HSP proteins are involved in response to environmental stress, it is possible these mutations play a role in high-elevation adaptation. In this study, I investigated the functional consequences of these substitutions and inferred their potential links to adaptation. I examined HSP90AA1 sequences of 13 anuran species previously studied. Using PROVEAN, I isolated three deleterious mutations: P65S, K195A, and _199I, each shared between two of the high-elevation species. I further analyzed the protein structure, stability change, and structural damage using model predictions. Based on its buried location and cavity expansion, P65S was predicted to most likely alter protein function. Furthermore, I examined HSP90AA1 sequences of over 100 other animal species available from public databases and found that serine at site 65 is ubiquitously present in cold-water fish, suggesting the substitution is related to cold adaptation. Alanine at site 195 and isoleucine at site 199 were not found in any other species, but these substitutions also might impact protein function as they are predicted to be destabilizing and their ancestral residues have reported post-translational modifications in orthologs. Tests of protein function and an investigation of more sequences from high-elevation species would help to further link these substitutions to adaptation, particularly P65S. Identifying mutations that contribute to high-elevation adaptation would aid in uncovering the molecular mechanisms of adaptation.

Keywords: molecular convergence, high-elevation adaptation, amphibian, HSP90AA1 gene, amino acid substitution

Introduction

While phenotypic convergences are often evident as traits that allow species to survive in a common environment, it is unclear whether these are caused by shared mutational paths (molecular convergence), or different ones. High-elevation adaptation is a prime candidate for investigating molecular convergent evolution, due to the physiological stress caused by low oxygen, cold temperatures, and high UV radiation (Cheviron & Brumfield, 2012). There is varied evidence regarding whether convergent and parallel amino acid substitutions, a major form of molecular convergent evolution, contribute to adaptation (Natarajan et al., 2015, Lu et al., 2020). Convergent substitutions describe different amino acids at a given genetic site in different species being replaced by the same amino acid in both lineages, whereas parallel substitutions result from the same ancestral residue. In

this report, “convergent substitutions” may be used generally to refer to both convergent and parallel substitutions.

Lu et al. (2020) detected a substantial amount of convergent and parallel amino acid substitutions in several genes of four high-elevation anurans (*Bufo tibetanus*, *Scutiger boulengeri*, *Rana kukunoris*, and *Nanorana parkeri*). These frogs are able to thrive at remarkable heights of around 3000 m in Tibet and are presumed to have evolved separately from low-elevation ancestors (Lu et al., 2020). This makes them compelling candidates to investigate the potential of molecular convergence linked to high-altitude adaptation. A particularly interesting convergent gene identified by Lu et al. (2020) is HSP90AA1, which encodes a heat shock protein. All four of the species shared at least one amino acid substitution with one of the other species in this protein, and an impressive total of 16 parallel and convergent substitutions were found

between *R. kukunoris* and *N. parkeri*. Heat shock proteins (HSPs) have many roles and are categorized by molecular weight. HSP90s in particular are important for maintaining homeostasis in the cell (Jackson, 2013). Their expression is associated with environmental stresses, including heat and cold stress. HSP90 proteins are molecular chaperones and are linked to ATPase activity across a range of species. In addition to being a convergent gene, Lu et al. (2020) also identified the HSP90AA1 gene as fast-evolving. These combined with its known functions suggest that the gene plays a role in adaptation to the challenges of high-elevation environments.

Functional analysis of the convergent substitutions in the HSP90AA1 gene of these anurans may provide further link to adaptation. It is predicted that the majority of convergent substitutions are likely due to chance and do not affect protein function or contribute to adaptation (Thomas and Hahn, 2015). Empirical studies have also demonstrated that many convergent substitutions with signals of positive selection are functionally inconsequential, such as in the case of Andean high-elevation waterfowls (Natarajan et al., 2015). Nevertheless, the high level of convergence of the HSP90AA1 gene among the high-elevation anurans studied by Lu et al. (2020), especially between two of the species, is notable and warrants further analysis.

In this study, I performed my own assessment to isolate convergent amino acid substitutions in the HSP90AA1 gene of the four anurans and further evaluated the substitutions using several bioinformatic tools, such as PROVEAN and Missense3D. The objective was to detect which substitutions are likely to have a functional impact, and what that impact may be in relation to high-elevation adaptation. Results from this study will provide foundation for future empirical functional verification experiments.

Methods

This study includes the 13 anuran species previously studied by Lu et al. (2020): four high-elevation species along with two close low-elevation relatives for each, and *Xenopus tropicalis* as the outgroup. The phylogeny from Lu et al. (2020) was used (Figure 1). All sequences were obtained from public databases (see Supplementary Information S1 for accession numbers).

The amino acid sequences in this analysis were aligned by Lu et al. (2020) using Clustal Omega and are partial, starting from “YIDQEE” and ending with “TLVSVT” in the sequence of *X. tropicalis*. Residues in this report are numbered accordingly from 1-262 (equivalent to 284-545 in UniprotKB – P07900 HS90A_HUMAN). I examined the amino acid composition of the sequences using a breakdown of amino acid residues from PredictProtein (Yachdav et al., 2014) to detect patterns of change between high-elevation species as compared to their low-elevation relatives. As nine out of 13 of the sequences were missing “YIDQEE” at positions 1-6 and two were missing “T” in the last position (262), I excluded

these seven sites from the analysis of amino acid composition and only assessed sites 7-261.

Convergent and parallel substitutions among high-elevation species were deduced by examining sequence differences between each set of a high-elevation species and its two low-elevation relatives. The effective substitutions were determined to be the amino acids in each high-elevation species that were different from the residue at the same site in both of its low-elevation close relatives. I used several approaches to examine the potential impact of detected substitutions on protein structure and function. Firstly, the substitutions were analyzed for effect with PROVEAN (Choi et al., 2012). I used PROVEAN results to narrow down substitutions that are likely functionally impactful to further analyze. Secondly, I examined several structural properties of these sites to see if these properties were affected by the mutated residues. I assessed location within the 3D structure as given by PremPS (Chen et al., 2020), as well as secondary structure and solvent accessibility predictions for these sites from I-TASSER (Yang et al., 2015). Thirdly, as loss of protein stability is known to be associated with disease and can also lead to the evolution of new enzymatic functions (Tokuriki et al., 2008), I investigated the predicted stability change caused by the substitutions of interest. A consensus stability prediction was made using unfolding free energy ($\Delta\Delta G$) as predicted by PremPS, folding free energy as predicted by DynaMut (Rodrigues et al., 2018), and vibrational entropy difference ($\Delta\Delta S_{Vib}$ ENCoM, signifying change in flexibility) also predicted by DynaMut. In addition, I used Missense3D (Ittisoponpisan et al., 2019), a tool which detects a variety of structural damages, to see if any damages could be predicted for these substitutions. 3D model files supplied by I-TASSER of the closest low-elevation relatives were used in all tools requiring a 3D model input (PremPS, DynaMut, Missense3D). The convergent substitution _199I was analyzed using both the sequences for *B. gargarizans* and *O. popei*, effectively testing both the S to I and T to I mutations. The sequence of *N. yunnanensis* was used for T199S, and the sequence of *R. chensinensis* was used for the remainder of the substitutions. Fourthly, I examined reported post-translational modifications at the sites of interest annotated by 3DBIONOTES (Segura et al., 2019) with data imported from PhosphositePlus (Hornbeck et al., 2015), as the presence of modifications could signify importance of conservation of these residues.

I also searched public databases, including Ensembl (Yates et al., 2020), UniProtKB (The UniProt Consortium, 2018), ICGC (International Cancer Genome Consortium, 2020), Pharos (Nguyen et al., 2017), and webtool SNPs3D (Yue et al., 2006), for reported mutations in the HSP90AA1 gene and associated impacts, to see if any matched the substitutions of interest. Knowing whether other animals have the same mutations, particularly other species at high altitudes, could be insightful. Hence, I also examined sequences from other species obtained from NCBI databases (NCBI Resource

Coordinators, 2018) for similar mutations at the sites of interest.

Results

Amino acid composition

The amino acid composition of the HSP90AA1 sequences of the four high-elevation species was compared to their closest low-elevation relatives (Figure 2). When examining individual amino acids, the only pattern common to all four species was a decrease in valine. In three out of the four species (*S. boulengeri*, *R. kukunoris*, and *N. parkeri*), there was an increase in phenylalanine and serine, and a decrease in asparagine. In three out of the four species (*B. tibetanus*, *R. kukunoris*, and *N. parkeri*), there was an increase in histidine and a decrease in proline. An evaluation of amino acids by their properties (polar/nonpolar, basic/acidic, aromatics) revealed no differences common to all four high-elevation species nor any differences common to three out of four.

Analysis of amino acid substitutions

A total of 61 effective single amino acid mutations were identified and analyzed with PROVEAN (Table 1). Among all effective substitutions of the four high-elevation species, serine was the most common target of mutation, being the target of nine mutations across seven different sites. This was not a surprise, considering serine has been found to be the most targetable of amino acids for substitutions (Creixell et al., 2012). Lysine was the most mutated, being replaced nine times at six sites. The mutation lysine to arginine accounted for five of those and was the second most common specific mutation. The reverse, arginine to lysine, was third most common and occurred four times at two unique sites. The mutation glutamate to aspartate was most common, occurring seven times at five sites. These mutations had neutral PROVEAN predictions, and this was not surprising considering the similar properties of glutamate and aspartate and likewise, lysine and arginine.

A total of 18 convergent/parallel mutations were identified. The majority of these substitutions were predicted to be neutral (Table 1). The three substitutions predicted to be deleterious (P65S, K195A, and _199I) were further analyzed, along with three neutral substitutions (_199S, E254D, and V259L; a total of five sites) for comparison and to help with determining the significance of bioinformatic results. Notably, all four high-elevation species exhibited a convergent substitution at site 199 with one other species.

Location, secondary structure, and solvent accessibility of mutated sites

PremPS determined site 65 to be in the core of the protein, with the other four sites located on the surface. The

secondary structure was predicted to be coil at sites 65 and 254, helix at sites 195 and 199, and strand at site 259. I-TASSER did not detect any differences in secondary structure between the high-elevation species and their low-elevation close relatives. The only difference in solvent accessibility predicted by I-TASSER (ranked from 0 to 9, 0 being buried and 9 being highly exposed) was an increase from 6 to 7 with mutation E254D. The solvent accessibilities for the other sites were 0 at site 65, 4 at site 259, 5 at site 195, and 7 at site 199.

Stability change and structural damage prediction

The consensus predictions using PremPS and DynaMut were stabilizing for M199S and destabilizing for the five other mutations analyzed. Damage was detected by Missense3D only in the mutation P65S. Specifically, it detected an expansion of cavity volume by 185.976 Å³. Using the sequence of *N. yunnanensis* instead of *R. chensinensis*, there was an expansion of cavity volume by 96.336 Å³.

Post-translational modifications (PTMs)

A search for reported PTMs of the HSP90AA1 gene revealed a few modifications at the sites of interest. Acetylation and ubiquitination were both reported at site 195 in humans, modifications which are exclusive to lysine residues. Ubiquitination at site 195 has also been reported in mice. At site 199 phosphorylation was reported, which can occur on threonine, tyrosine, and serine. Four other convergent sites (72, 183, 198, and 216) were also found to have at least one PTM in humans.

Reported mutations of HSP90AA1 in humans and other species

Some of the same mutations detected in the high-elevation frogs have also been reported in other organisms, particularly humans for which the most variant data exists, but functional significance remains unclear. From the online resources searched, no variants were found to exactly match the deleterious mutations studied. However, a lysine to glutamine mutation has been reported at site 195 in humans, as well as a threonine to proline mutation at site 199 (variant IDs rs923000772 and rs1361885338 respectively, Ensembl release 100). No clinical significance or functional consequences have been reported for these mutations, but they both have a predicted “medium” impact based on Mutation Assessor.

HSP90AA1 sequences of 107 species from a range of animal groups were retrieved from public databases. The vast majority had proline, lysine, and threonine at the sites 65, 195, and 199 respectively. This matched the outgroup in this study, as well as five out of the eight low-elevation species (Table 2). The 107 species were categorized into low elevation (<1500 m), high elevation (primarily >1500 m), mixed

elevation for those occurring at both low and high elevations, and unknown, based on the consensus result of a web search. In the twelve species categorized as high-elevation species (with five of these being found mainly above 2500 m), all had proline, lysine, and threonine at sites 65, 195, and 199 respectively.

Out of the three sites, lysine at site 195 was the most conserved. Only one species had a residue other than lysine, which was serine in the American alligator (*Alligator mississippiensis*). Most of the differences at the other two sites occurred in fish. Some fish had phenylalanine in place of proline at site 65: the zebra mbuna (*Maylandia zebra*), spotted gar (*Lepisosteus oculatus*), zebrafish (*Danio rerio*), zig-zag eel (*Mastacembelus armatus*), Indian medaka (*Orzias melastigma*), and guppy (*Poecilia reticulata*). Some fish were found to have serine, matching the substitution in two of the high-elevation anurans. The remaining nine fish including three chondrichthyans, as well as all other species examined, had proline at site 65. The fish with serine notably live in cold waters, suggesting a link between serine at site 65 and environmental adaptation. One of them, the Chinook salmon (*Oncorhynchus tshawytscha*), also occurs at mixed elevations and is known to migrate up to 1500 m in altitude ("Chinook Salmon", n.d.). In light of this, I proceeded to target fish sequences in my analysis.

Eleven fish in total were found to have serine at site 65: the wolf eel (*Anarrhichthys ocellatus*), Atlantic cod (*Gadus morhua*), Chinook salmon, orangethroat darter (*Etheostoma spectabile*), lumpfish (*Cyclopterus lumpus*), emerald rockcod (*Trematomus bernacchii*), pinecone soldierfish (*Myripristis murdjan*), Pacific halibut (*Hippoglossus stenolepis*), New Zealand spotty (*Notolabrus celidotus*), Atlantic herring (*Clupea harengus*), and northern pike (*Esox lucius*). All prefer cooler waters (mean temperature <15°C) except for the pinecone soldierfish, which is found in the Indo-Pacific and thrives in water 23-26°C ("Pinecone Soldierfish", 2008). One fish, the emerald rockcod, is an Antarctic bottom-dweller and lives at a temperature of -1.86°C. Out of the 15 fish with residues other than serine at site 65, only one, the thorny skate (*Amblyraja radiata*), lives in cold water.

A few variations were found at site 199. Many fish (18/26) had serine, but the sailfin molly (*Poecilia latipinna*), Indian glassy fish (*Parambassis ranga*), coelacanth (*Latimeria chalumnae*), milkfish (*Chanos chanos*), Mexican tetra (*Astyanax mexicanus*), and orangethroat darter had threonine. One fish, the Japanese grenadier anchovy (*Coilia nasus*), had alanine, and the New Zealand spotty had cysteine. The American alligator also had serine at site 199, as well as the three invertebrates examined (*Hydra vulgaris*, *Trichonephila clavipes*, *Bugula neritina*). Threonine was found in all other animals besides the three marsupials, which had methionine.

Discussion

The high-elevation species shared a few amino acid composition biases, which may have important implications. In the case of *R. kukunoris* and *N. parkeri*, it is likely the shared patterns are a result of the high number of parallel substitutions between them, but it is interesting that they shared these patterns with the other high-elevation species despite not sharing the same substitutions. Serine, asparagine, and valine have been found to have a high degree of mutability relative to other amino acids, while histidine and proline are moderate, and phenylalanine is low (Creixell et al., 2012). Thus, the shared increase in phenylalanine, an uncommon target of mutation, is more interesting, especially considering the substitutions involving phenylalanine are at different sites in each species. The decrease in valine among all four species is also interesting, as valine is more likely to be the target of mutation than to be mutated (Creixell et al., 2012). More research about shared amino acid changes is needed to explore the possible connotations of these patterns, if any.

Not surprisingly, the majority of detected convergent mutations likely do not have any impact on protein function. However, three substitutions at sites 65, 195, and 199 were predicted to be deleterious. Based on location and secondary structure, the mutation P65S stands out and likely has a high impact. It has been observed that buried mutations with low solvent accessibility have greater ability to affect protein stability and function than mutations at the surface (Ramsey et al., 2011). Similarly, Tokuriki et al. (2008) found that mutations that confer new functions are more likely to be in the core and part of a random coil than other regions. Site 65 resides in the core of the protein, has low solvent accessibility, and is part of a coil. Moreover, P65S is the only examined mutation to have predicted structural damage by Missense3D, specifically, an increase in cavity volume. Cavities are associated with a loss of stability and yet, the greater flexibility they provide to the protein seems to have functional importance (Vallone & Brunori, 2004). Cavities allow for different conformational changes, so an increase in cavity volume may lead to new conformations and in turn impact binding and function. It has been experimentally found that cavity-creating mutations usually result in surrounding atoms slightly adjusting their side chains to reduce the cavity (Machicado et al., 2002; Eriksson et al., 1992), but functional impact is uncertain. The details surrounding the increased cavity volume of the P65S mutation remain to be examined.

Aside from structural change predictions, the reported PTMs at sites 195 and 199 are additional evidence that these sites are important to protein function. Phosphorylated residues have been found to be more conserved than non-phosphorylated residues, especially serine and threonine, indicating these residues have an important functional role (Creixell et al., 2012). Assuming anurans also possess these PTMs at sites 195 and 199, it is likely that substitutions at these sites will impact function when the replacement is

unable to be modified in the same way. This applies to K195A, T199I, and S199I.

While the deleterious substitutions have not been reported in other studies, an examination of HSP90AA1 sequences from other species revealed the serine substitution at site 65 in several fish species. Considering that ten out of eleven fish with serine at site 65 live in cold water, compared to only one out of 15 fish with residues other than serine, there is strong evidence this substitution is related to cold adaptation. The _199I and K195A substitutions have not been reported in other species. Since many fish and the invertebrates had serine at site 199, it is possible that serine was in fact the ancestral amino acid at this site instead of threonine, and this interpretation would require just one more mutation event (nine rather than eight) to account for the variety in the 13 anurans. The fact that only one species in my search had a residue other than lysine at site 195 stands out considering four of the 13 anurans had a residue other than lysine. Overall, it seems there is a high amount of variation among the frogs in the study compared to other species examined at the sites of interest, which could signify fast evolution of this gene in these anurans. However, this could also be due to lack in number or diversity of samples. It may be insightful to investigate sequences of more high-elevation species across the animal kingdom along with their low-elevation close relatives.

Elevation categories for other species in the HSP90AA1 search were approximate. It is possible that some sequences were taken from populations at different elevations than average for the species, leading to a miscategorization. It is also important to note that a general level of caution should be exercised when using bioinformatic tools. In addition, the programs used assessed the effects of single mutations. It is possible that mutations that have no effect on their own could have an effect when combined, and vice versa. Hence, I used several different methods to assess impact of the mutations. Considering overall results, serine at site 65 is an excellent candidate for functional verification for cold adaptation.

In summary, the HSP90AA1 gene exhibits convergent evolution among high-elevation anurans and could be linked to high-elevation adaptation. While most of the shared amino acid substitutions are likely inconsequential, I identified and analyzed three substitutions that are predicted to be functionally impactful. P65S is particularly interesting due to its structural location and its predicted effect on cavity volume. This substitution is highly favoured in cold-water fish, providing evidence that it is involved in environmental adaptation. The other two substitutions, K195A and _199I, have not been reported in other species but may also have an effect based on other methods of analysis. Experimental tests are needed to confirm and further explore any change to the protein these substitutions cause and how they may contribute to high-elevation adaptation.

Acknowledgements

This study is supported by an Anne Innis Dagg Summer Research Assistantship to the author; thank you to Dr. Anne Dagg and the College of Biological Sciences for the scholarship. Thank you to Dr. Steffen Graether for his insight regarding protein stability change and the use of predictive tools. Finally, thank you to Dr. J. Fu for his endorsement, his supervision over the entire study, and his edits and encouragement throughout the writing of this report

References

- Chen, Y., Lu, H., Zhang, N., Zhu, Z., Wang, S., & Li, M. (2020). PremPS: Predicting the effects of single mutations on protein stability. *bioRxiv*, 029074. doi:10.1101/2020.04.07.029074
- Cheviron, Z., & Brumfield, R. (2012). Genomic insights into adaptation to high-altitude environments. *Heredity*, 108, 354–361. doi:10.1038/hdy.2011.85
- Chinook Salmon. (n.d.). Retrieved from <https://oceana.org/marine-life/ocean-fishes/chinook-salmon>
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE*, 7(10), e46688. doi:10.1371/journal.pone.0046688
- Creixell, P., Schoof, E. M., Tan, C. S. H., & Linding, R. (2012). Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues. *Philosophical Transactions of the Royal Society. B*, 367(1602), 2584–2593. doi: 10.1098/rstb.2012.0076
- Eriksson, A., Baase, W., Zhang, X., Heinz, D., Blaber, M., Baldwin, E., & Matthews, B. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, 255, 178–183. doi:10.1126/science.1553543
- Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, 43, D512–D520. doi:10.1093/nar/gku1267
- International Cancer Genome Consortium. (2020). ICGC Data Portal 6.0.11. <https://dcc.icgc.org/>
- Ittisoponpisan, S., Islam, S. A., Khanna, T., Alhuzimi, E., David, A., & Sternberg, M. J. E. (2019). Can predicted protein 3D structures provide reliable insights into whether missense variants Are disease associated? *Journal of Molecular Biology*, 431, 2197–2212. doi:10.1016/j.jmb.2019.04.009

Analysis of convergent and parallel amino acid substitutions in the HSP90AA1 gene among high-elevation anurans (Tao)

- Jackson, S. E. (2013). Hsp90: structure and function. *Top Current Chemistry*, 328, 155-240. doi: 10.1007/128_2012_356
- Lu, B., Jin, H., & Fu, J. (2020). Molecular convergent and parallel evolution among four high-elevation anuran species from the Tibetan region. *BMC Genomics*, 21, 839. doi:10.1186/s12864-020-07269-4
- Machicado, C., Bueno, M., & Sancho, J. (2002). Predicting the structure of protein cavities created by mutation. *Protein Engineering, Design and Selection*, 15(8), 669-675. doi:10.1093/protein/15.8.669
- Natarajan, C., Projecto-Garcia, J., Moriyama, H., Weber, R. E., Muñoz-Fuentes, V., et al. (2015). Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves limited parallelism at the molecular sequence level. *PLOS Genetics*, 11(12), e1005681. doi:10.1371/journal.pgen.1005681
- NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 46(D1), D8-D13. doi:10.1093/nar/gkx1095
- Nguyen, D-T., Mathias, S. et al. (2017). Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Research*, 45(D1), D995-D1002. doi:10.1093/nar/gkw1072
- Pinecone Soldierfish. (2008). Retrieved from <https://saltwater.aqua-fish.net/?pinecone-soldierfish>
- Ramsey, D. C., Scherrer, M. P., Zhou, T., & Wilke, C. O. (2011). The relationship between relative solvent accessibility and evolutionary rate in protein evolution. *Genetics*, 188(2), 479-488. doi:10.1534/genetics.111.128025
- Rodrigues, C. H. M., Pires, D. E. V., & Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, 46(W1), W350-W355. doi:10.1093/nar/gky300
- Segura, J., Sanchez-Garcia, R., Sorzano, C. O. S., & Carazo, J. M. (2019). 3DBIONOTES v3.0: Crossing molecular and structural biology data with genomic variations. *Bioinformatics*, 35(18), 3512-3513. doi:10.1093/bioinformatics/btz118
- The UniProt Consortium. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 46(5), 2699 (2018). doi:10.1093/nar/gky092
- Thomas, G. W., & Hahn, M. W. (2015). Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Molecular Biology and Evolution*, 32, 1232-1236.
- Tokuriki, N., Stricher, F., Serrano, L., & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLOS Computational Biology*, 4(2), e1000002. doi:10.1371/journal.pcbi.1000002
- Vallone, B., & Brunori, M. (2004). Roles for holes: are cavities in proteins mere packing defects. *Italian Journal of Biochemistry*, 53(1), 46-52. Retrieved from <http://researchgate.net>
- Yachdav, G., Koppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., et al. (2014). PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Research*, 42(W1), W337-W343. doi:10.1093/nar/gku366
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, 12, 7-8. doi:10.1038/nprot.2010.5
- Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. *Nucleic Acids Research*, 48(D1), D682-D688. doi:10.1093/nar/gkz966
- Yue, P., Melamud, E., & Moul, J. (2006). SNPs3D: Candidate gene and SNP selection for association studies. *BMC Bioinformatics*, 7(166). doi:10.1186/1471-2105-7-166

Analysis of convergent and parallel amino acid substitutions in the HSP90AA1 gene among high-elevation anurans (Tao)

Figure 2. Amino acid variations between each of the high-elevation anurans (shown in bold) and their closest low-elevation relative in this study. Only sites with differing amino acids in each comparison are shown. Amino acid position is approximate and one-letter codes are displayed.

Table 1. Amino acid substitutions in four high-elevation anurans when compared to low-elevation relatives. Convergent/parallel substitutions i.e. those shared between two or more species are marked with an asterisk.

Species compared	# of amino acid differences	# of common differences	Effective substitutions	PROVEAN prediction
<i>B. tibetanus</i> vs. <i>B. gargarizans</i>	32	5	E24D	Neutral (-1.721)
			P61A	Deleterious (-6.371)
			162I	Neutral (-0.625)
<i>B. tibetanus</i> vs. <i>R. marina</i>	6	5	199I*	Deleterious (-2.833)
			V259L*	Neutral (-1.935)
<i>S. boulengeri</i> vs. <i>O. popei</i>	12	5	K148Q	Deleterious (-2.546)
			N172S	Deleterious (-3.980)
			L176F	Deleterious (-2.530)
<i>S. boulengeri</i> vs. <i>L. boringii</i>	26	5	199I*	Deleterious (-3.099)
			K202P	Deleterious (-3.297)
<i>R. kukunoris</i> vs. <i>R. chensinensis</i>	27	26	T34S	Neutral (-1.217)
			A64T	Deleterious (-2.821)
			P65S*	Deleterious (-6.547)
			R72K*	Neutral (0.290)
			K73R*	Neutral (-2.398)
			E93D*	Neutral (-0.700)
			Y98F	Neutral (-2.203)
			R103K*	Neutral (1.266)
			M119T	Neutral (0.328)
			L144M*	Neutral (-1.086)
			S145A	Neutral (1.061)
			E149D	Neutral (-0.490)

Analysis of convergent and parallel amino acid substitutions in the HSP90AA1 gene among high-elevation anurans (Tao)

<i>R. kukunoris</i> vs. <i>O. margaretae</i>	29		S170T	Neutral (-1.662)
			N174K*	Neutral (0.801)
			Y183H*	Neutral (1.344)
			V192I*	Neutral (-0.450)
			K195A*	Deleterious (-3.931)
			C198V*	Neutral (3.285)
			I199S*	Neutral (2.593)
			K216R*	Neutral (-1.111)
			E217D	Neutral (-0.198)
			I231H	Neutral (1.967)
			I236L*	Neutral (0.078)
			I239T*	Neutral (0.422)
			E254D*	Neutral (1.316)
			V259L*	Neutral (-1.836)
<i>N. parkeri</i> vs. <i>N. yunnanensis</i>	26	25	T22S	Neutral (-1.407)
			V60L	Neutral (-1.654)
			P65S*	Deleterious (-6.325)
			R72K*	Neutral (0.306)
			K73R*	Neutral (-2.315)
			K75R	Neutral (-0.343)
			N90S	Deleterious (-2.918)
			E93D*	Neutral (-0.188)
			N100S	Neutral (-2.114)
			R103K*	Neutral (1.116)
			L144M*	Neutral (-1.064)
			L154F	Neutral (4.178)
			S170N	Neutral (-1.605)
<i>N. parkeri</i> vs. <i>Q. spinosa</i>	26		N174K*	Neutral (0.951)
			L179M	Neutral (-1.241)
			Y183H*	Neutral (1.677)
			M191T	Deleterious (-2.895)
			V192I*	Neutral (-0.384)
			K195A*	Deleterious (-3.776)
			C198V*	Neutral (3.124)
			T199S*	Neutral (-0.085)
			K216R*	Neutral (-1.055)
			I236L*	Neutral (0.045)
			I239T*	Neutral (-0.161)
			E254D*	Neutral (1.249)

Analysis of convergent and parallel amino acid substitutions in the HSP90AA1 gene among high-elevation anurans (Tao)

Table 2. Amino acid present at the three convergent sites with predicted deleterious mutations across the thirteen anuran species studied. High-elevation species are in bold.

Species	Site 65	Site 195	Site 199
<i>Bufo tibetanus</i>	P	K	I
<i>Bufo gargarizans</i>	P	T	S
<i>Rhinella marina</i>	P	K	T
<i>Nanorana parkeri</i>	S	A	S
<i>Nanorana yunnanensis</i>	P	K	T
<i>Quasipaa spinosa</i>	P	K	T
<i>Rana kukunoris</i>	S	A	S
<i>Rana chensinensis</i>	P	K	M
<i>Odorrana margaretae</i>	P	K	T
<i>Scutiger boulengeri</i>	P	K	I
<i>Oreolalax popei</i>	P	K	T
<i>Leptobrachium boringii</i>	T	S	S
<i>Xenopus tropicalis</i>	P	K	T
Most common residue in 107 other animals	P	K	T

Supplementary Information

S1. Accession numbers for the HSP90AA1 gene sequences of the anurans used in this study.

1. *Rhinella marina*, NCBI assembly accession 436 GCA 900303285.1
2. *Xenopus tropicalis*, NCBI assembly accession GCF 000004195.
3. *Nanorana parkeri*, NCBI assembly accession GCF 000935625.1
4. *Bufo gargarizans*, NCBI BioProject PRJNA383934
5. *Leptobrachium boringii*, NCBI Gene Expression Omnibus GSE89016
6. *Nanorana yunnanensis*, BIGD BioProject 440 PRJCA000409
7. *Odorrana margaretae*, NCBI Sequence Reads Archive SRA091981
8. *Rana chensinensis* and *Rana kukunoris*, NCBI Sequence Reads Archive SRA060325
9. *Bufo tibetanus*, *Quasipaa spinosa*, and *Scutiger boulengeri*, NCBI accession number PRJNA524747
10. *Oreolalax popei*, accession number PRJNA357944