*Article*

# Concerted and Independent Evolution of Control Regions 1 and 2 of Water Monitor Lizards (*Varanus salvator macromaculatus*) and Different Phylogenetic Informative Markers

Watcharaporn Thapana [1,2,†], Nattakan Ariyaraphong [1,2,3,†], Parinya Wongtienchai [2,†], Nararat Laopichienpong [1,3], Worapong Singchat [1,3], Thitipong Panthum [1,3], Syed Farhan Ahmad [1,3], Ekaphan Kraichak [4], Narongrit Muangmai [5], Prateep Duengkae [1,3] and Kornsorn Srikulnath [1,2,3,6,*]

1. Animal Genomics and Bioresource Research Center (AGB Research Center), Faculty of Science, Kasetsart University, 50 Ngamwongwan, Bangkok 10900, Thailand; golf_foureye@hotmail.com (W.T.); nattakan.ari@ku.th (N.A.); nararat.l@ku.th (N.L.); fsciwos@ku.ac.th (W.S.); thitipong.pa@ku.th (T.P.); syedfarhan.a@ku.th (S.F.A.); fforptd@ku.ac.th (P.D.)
2. Laboratory of Animal Cytogenetics and Comparative Genomics (ACCG), Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngamwongwan, Bangkok 10900, Thailand; parinya_wongtienchai@hotmail.com
3. Special Research Unit for Wildlife Genomics (SRUWG), Department of Forest Biology, Faculty of Forestry, Kasetsart University, 50 Ngamwongwan, Bangkok 10900, Thailand
4. Department of Botany, Faculty of Science, Kasetsart University, 50 Ngamwongwan, Bangkok 10900, Thailand; ekaphan.k@ku.th
5. Department of Fishery Biology, Faculty of Fisheries, Kasetsart University, 50 Ngamwongwan, Bangkok 10900, Thailand; ffisnrm@ku.ac.th
6. Amphibian Research Center, Hiroshima University, 1-3-1, Kagamiyama, Higashihiroshima 739-8526, Japan
* Correspondence: kornsorn.s@ku.ac.th; Tel.: +662-562-5555; Fax: +662-942-8290
† These authors contributed equally to this work.

**Simple Summary:** The evolutionary patterns and phylogenetic utility of duplicate control regions (CRs) in 72 individuals of *Varanus salvator macromaculatus* and other varanids have been observed. Divergence of the two CRs from each individual revealed a pattern of independent evolution in CRs of varanid lineage. This study is a first step towards developing new phylogenetic evolutionary models of the varanid lineage, with accurate evolutionary inferences to provide basic insights into the biology of mitogenomes.

**Abstract:** Duplicate control regions (CRs) have been observed in the mitochondrial genomes (mitogenomes) of most varanids. Duplicate CRs have evolved in either concerted or independent evolution in vertebrates, but whether an evolutionary pattern exists in varanids remains unknown. Therefore, we conducted this study to analyze the evolutionary patterns and phylogenetic utilities of duplicate CRs in 72 individuals of *Varanus salvator macromaculatus* and other varanids. Sequence analyses and phylogenetic relationships revealed that divergence between orthologous copies from different individuals was lower than in paralogous copies from the same individual, suggesting an independent evolution of the two CRs. Distinct trees and recombination testing derived from CR1 and CR2 suggested that recombination events occurred between CRs during the evolutionary process. A comparison of substitution saturation showed the potential of CR2 as a phylogenetic marker. By contrast, duplicate CRs of the four examined varanids had similar sequences within species, suggesting typical characteristics of concerted evolution. The results provide a better understanding of the molecular evolutionary processes related to the mitogenomes of the varanid lineage.

**Keywords:** varanid; control region; ortholog; paralog

## 1. Introduction

The mitochondrial control region (mtCR) is a major noncoding segment of the vertebrate mitochondrial genome (mitogenome). The region includes the displacement loop (D-loop), which comprises the third strand of DNA, thus creating a semi-stable structure [1]. The mtCR plays an important role in transcriptional and translational regulation of protein-coding sequences, or it serves as the origin of DNA replication [2]. The nucleotide CR sequence is the most rapidly evolving region of the mitogenome, and it lacks coding sequences; thus, it is widely used as a molecular marker in population genetics, phylogenetic studies, and phylogeographic studies [3–5]. Vertebrates such as birds, snakes, turtles, and fish exhibit segmental duplications within the CR or an entire duplication of the CR, leading to the formation of repeats or possible homogenization between the duplicated copies of CR [6–8].

GenBank contains 17,489 complete mitogenomes for squamate reptiles (as of April 2021, http://www.ncbi.nlm.nih.gov/genome), with several duplicate CRs observed in varanids and snakes [6,7]. A comparison between two CRs (CR1 and CR2) revealed identical or highly similar nucleotide sequences, similar to the concerted evolution as found in Bothidae and Samaridae [9]. By contrast, orthologous copies of duplicate CRs from different species such as in varanid, gecko lizard, and platysternid lineages, are genetically closer to each other than to paralogous copies of duplicate CRs (CR1 and CR2) within the same species [6]. This might be a result of the independent evolution of the two copies after an ancient duplication event, although the mechanism behind such an event is not clearly understood [8].

Varanids or monitor lizards comprise a single extant genus, *Varanus*, within the family Varanidae. To date, around 80 extant species have been described and distributed in Afro-Arabia, Western to Southeast Asia, the Indonesian Archipelago, Papua New Guinea, and Australia [10]. Mitogenomes of the Komodo dragon (*V. komodoensis*; Ouwens 1912) [11,12] and Nile monitor (*V. niloticus*; Linnaeus 1758) [13,14] have unique gene organization features. Genes between the NADH dehydrogenase subunit 6 (*ND6*) gene and proline tRNA gene are extensively shuffled, and the CR has been duplicated in an ancestral varanid lineage during the Paleocene age or earlier [15]. This is consistent with the Cenozoic over-water dispersal of Southeast Asian varanids, such as the water monitor (*V. salvator macromaculatus*; Deraniyagala 1944) [16] across the Indonesian Archipelago and Komodo dragon (*V. komodoensis*) [15]. The presence of duplicate CRs in varanid mitogenomes is an intriguing structural phenomenon and raises basic questions concerning how the nucleotide sequences of duplicate CRs remained similar over time. Variations in CRs at the population and species level in varanids have not been fully elucidated [5,17]. In light of this scenario, we propose two hypotheses: (1) orthologous copies of duplicate CRs in different individuals are genetically similar due to independent evolution, or (2) two CRs (CR1 and CR2) as paralogous copies exhibit identical or highly similar nucleotide sequences from concerted evolution. To characterize the variations in varanid CRs, we conducted this study to analyze the CR sequences of four varanids, namely *V. salvator* (*V. salvator macromaculatus* and *V. salvator komaini*), *V. exanthematicus*, *V. komodoensis* and *V. niloticus*, and 72 water monitors (*V. salvator macromaculatus*) (Table 1). We also compared and analyzed the sequence variations of mtCRs. These analyses have important implications in the selection of priority mitochondrial regions to assess the evolution and genetic diversity of varanid populations.

## 2. Materials and Methods

### 2.1. Specimen Collection and DNA Extraction

Blood specimens of water monitors (*V. salvator macromaculatus*) were collected from the ventral tail vein using a 23-gauge needle attached to a 2 mL disposable syringe containing 10 mM ethylenediaminetetraacetic acid for DNA extraction as previously reported by Wong-tienchai et al. [5] (Supplementary Table S1). Samples were collected from 47 individuals at the Bang Kachao Peninsula, Samut Prakan, (13°59′2″ N, 99°59′38″ E) and from 25 individ-

uals at Varanus Farm Kamphaeng Saen, Nakhon Pathom (14°00′59.9″ N, 99°57′46.8″ E). Permission was granted by the Sri Nakhon Khuean Khan Park (Royal Forest Department, Ministry of Natural Resources and Environment) and Kasetsart University (0909.6/15779). All experimental procedures involving animals conformed to the guidelines established by the Animal Care Committee of Kasetsart University, Thailand. Total genomic DNA was extracted according to the standard salting-out protocol, as previously described [18]. DNA quality and concentration were determined using 1% agarose gel electrophoresis and spectrophotometry.

*2.2. CR1 Sequencing*

The positions of duplicate CRs were determined from the locations of tRNA genes (tRNAPro and tRNAVal for CR1; tRNATyr and tRNASer for CR2) as the flanking regions of CR1 and CR2 in the mitogenome of *V. salvator* (GenBank accession number: EU747731). The CR1 fragments were amplified using the following primers: VSA_CR1 F (5′-ATTAATACCCAATTTTCCTTGCTC-3′) and VSA_CR1 R (5′-GCCCAGTGACCATTAATAT CAACT-3′), which were designed based on five varanid mtDNA sequences, namely *V. salvator macromaculatus* (GenBank accession number: AB980995), *V. salvator komaini* (GenBank accession number: AB980996), *Varanus exanthematicus* (GenBank accession number: AB738957), *Varanus komodoensis* (GenBank accession number: AB080276), and *Varanus niloticus* (GenBank accession number: AB185327). The positions of all primers were located in tRNA genes (tRNAPro and tRNAVal for CR1; tRNATyr and tRNASer for CR2) that are highly conserved along all varanus mitogenomes; therefore, entire sequences of both CRs were collected. Polymerase chain reaction (PCR) amplification was performed using 20 µL of 1× ThermoPol buffer containing 1.5 mM $MgCl_2$, 0.2 dNTPs, 5.0 µM primers, 0.5 U of *Taq* polymerase (Apsalagen Co., Ltd., Bangkok, Thailand), and 25 ng of genomic DNA. The PCR conditions were as follows: initial denaturation at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, 57 °C for 30 s, and 72 °C for 40 s, and a final extension at 72 °C for 5 min [5]. The PCR products were separated via electrophoresis on 1% agarose gels, and they were then cloned using the pGEM®-T Easy vector (Promega Corporation, Madison, WI, USA). Nucleotide sequences of DNA fragments were determined using the DNA sequencing service of First Base Laboratories Sdn Bhd (Seri Kembangan, Selangor, Malaysia). BLASTn programs (http://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to search nucleotide sequences in the National Center for Biotechnology Information database to confirm the identities of amplified DNA fragments. The generated sequences were deposited in the DNA Data Bank of Japan. The mitochondrial CR2 dataset used in our previous study was retrieved from the database [5].

*2.3. Positional Annotation*

Three functional regions, including the terminal-associated sequence (TAS), central conserved domain (CD), and conserved sequence blocks (CSB), were tentatively investigated in both CRs by recognizing sequences similar to those found in other vertebrates [3,6,12]. Variable number of tandem repeats (VNTRs) have been reported to exist only at the 3′ end of both CRs [19,20]. Tandem repeat sequences, including the motif, length of repeats, and copy number in the CR region, were investigated using the Tandem Repeats Finder 4.09 program [21].

*2.4. Comparison of Genetic Variability Based on CR Sequences at the Population Level*

Multiple sequence alignment was performed for 72 sequences in both CRs using the default parameters of Molecular Evolutionary Genetics Analysis X (MEGAX) software (Center for Evolutionary Functional Genomics, The Biodesign Institute, Tempe, PA, USA; [22]). Estimates of haplotype (*h*), nucleotide (π) diversity [23], and number of haplotypes (H) were calculated based on CR1 and CR2 sequences, as implemented in DnaSP version 6 [24]. A statistical parsimony network of consensus sequences was constructed using the Templeton, Crandall, and Sing (TCS) algorithm implemented in PopART version 1.7. to

address haplotype grouping [25]. The mitochondrial CR2 dataset employed in our previous study [5] was used for all analyses, similar to the CR1 dataset. The means and standard deviations of *h* and π diversity of both CRs were used to calculate t-statistics and *p*-values for two-sample *t*-test comparisons, following the formula in [26] in R version 4.0.3 [27].

### 2.5. Phylogenetic Analysis Based on Mitochondrial CR1 and CR2 Sequences at the Species Level

Substitution saturation decreases the amount of phylogenetic signal to the point that sequence similarities could be a result of chance alone rather than homology. Consequently, when saturation is achieved, the phylogenetic signal is lost, and the sequences no longer reveal the underlying evolutionary mechanisms [28]. The saturation of substitutions was evaluated by plotting the number of transitions (s) and transversions (v) against the K80 [29] sequence divergences as well as by comparing the information entropy-based index ($I_{ss}$) with critical values ($I_{ss.c}$) [30,31], as implemented in DAMBE7 [32]. If $I_{ss}$ is significantly lower than $I_{ss.c}$, the sequences do not experience substitution saturation. Phylogenetic analyses were performed using CR1 and CR2 datasets with the maximum likelihood (ML) reconstructed in IQ-TREE [33] using a model finder with the options TEST and –AICc, a tree search with 1000 bootstrap replicates, and Bayesian inference (BI) with MrBayes version 3.2.6 [34]. The GenBank database of four varanids is shown in Table 1 as of April 2021. The best-fit model of DNA substitution was determined for each CR using Kakusan4 [35]. The Markov chain Monte Carlo process was used to simultaneously run four chains for one million generations. After stabilization of the log-likelihood value, a sampling procedure was performed every 100 generations to obtain 10,000 trees, from which a majority-rule consensus tree with average branch lengths was generated. All sample points were discarded before attaining convergence as burn-in, and the Bayesian posterior probability in the sampled tree population was calculated as a percentage. The genetic distances of *p*-distance between CR sequences were calculated using the MEGAX program [22].

**Table 1.** Species used with accession numbers.

| Species | GenBank Accession Number | CRs | | Reference |
| --- | --- | --- | --- | --- |
| | | CR1 | CR2 | |
| *Varanus salvator macromaculatus* | LC326253-LC326324 | CR1 | - | This study |
| *Varanus salvator macromaculatus* | LC326325-LC326396 | - | CR2 | Wongtienchai et al. [5] |
| *Varanus salvator* | EU747731 | CR1 | CR2 | Castoe et al. [36] |
| *Varanus salvator macromaculatus* | AB980995 | CR1 | CR2 | Chaiprasertsri et al. [37] |
| *Varanus salvator komaini* | AB980996 | CR1 | CR2 | Chaiprasertsri et al. [37] |
| *Varanus exanthematicus* | AB738957 | CR1 | CR2 | - |
| *Varanus komodoensis* | AB080276 | CR1 | CR2 | Kumazawa and Endo [11] |
| *Varanus niloticus* | AB185327 | CR1 | CR2 | Kumazawa [13] |

### 2.6. Recombination Testing

Discordant evolutionary signals were detected when the phylogenetic trees were separately reconstructed from different regions. These conflicting signals are due to the recombination of duplicate CRs. To further analyze these signals, the following recombination tests were conducted for both the CRs: (1) Recombination Detection Program (RDP) [38], (2) Geneconv [39], (3) Maxchi [40], and (4) Chimaera [41]. These analyses were performed using the Recombination Detection Program, RDP5 [42], with previously described parameters [43]. All analyses were performed for all individuals to check for recombination occurrence in both CRs of water monitors.

## 3. Results

### 3.1. Positional Annotation in the Control Regions of V. salvator macromaculatus

Three conserved functional sections, including the TAS, CD, and CSB domains, were analyzed in both CR1 and CR2 of all the 72 individuals; however, no CD was observed in either CR. The TAS domain contained 78 bp for CR1 and 80 bp for CR2 in water monitor lizards. The conserved nucleotide sequence of TAS between CR1 and CR2 was 5′-TAGTT-3′. The CSB domain contained three conserved blocks, namely CSB-1 (5′-TTAATGGTCDCNGGRHAT -3′), CSB-2 (5′-DHWDBYMYNYHHDCYYYC -3′), and CSB-3 (5′-GCYHWDYRKTYAHMMAA-3′) for CR1, and CSB-1 (5′-TTCATYWYHAWWWWWTTBDN-3′), CSB-2 (5′-WWWWYCMYYWWHYYYY -3′), and CSB-3 (5′-GCYHWWYRKTYAHMAA -3′) for CR2. VNTRs were identified in CR1 (TCGCGCCACCTCCAGGATT), with two copies for one individual only, and CR2, (TTTTTTAAAAAAATTTTTTAT), (AAAAAAATTTTTTTA), (TTAAAAAAATTTTTT), and (AAAAAAATTTTTTATTTTTTTAA), ranging from to 2 to 4 copies in all individuals (Figure 1).
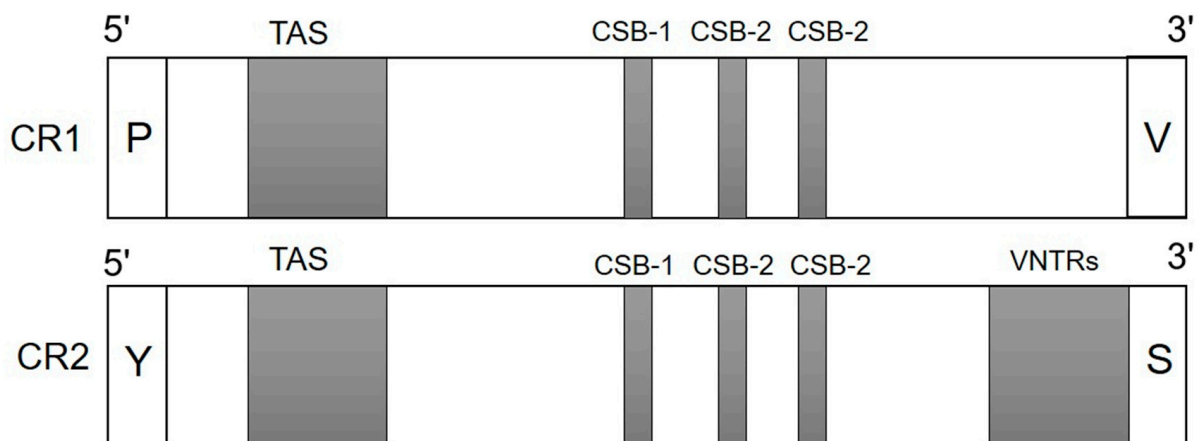


**Figure 1.** Structures of duplicate control regions (CRs) in all individuals of *Varanus salvator macromaculatus* (Deraniyagala 1944 [16]) in this study. Two functional regions, TAS and CSB, were detected in both the CRs of all individuals. The core sequences of these regions were found to be identical in both CR1 and CR2. Variable numbers of tandem repeats were detected only in CR2.

### 3.2. Sequence Variation in the CRs of V. salvator macromaculatus

The alignment lengths of CR1 and CR2 sequences were 663 and 867 bp, respectively. The number of haplotypes in CR1 was 44, and the number of haplotypes in CR2 was 52. The overall haplotype and nucleotide diversities were $0.935 \pm 0.019$ and $0.004 \pm 0.001$ for CR1 and $0.968 \pm 0.013$ and $0.004 \pm 0.001$ for CR2. Results of the *t*-test showed that the means of *h* and $\pi$ significantly differed between CRs (t = 3.974, df = 141.77, *p*-value < 0.01 for *h* value and t = 6.8043, df = 127.83, *p*-value < 0.01 for $\pi$ value). Meanwhile, results of the *t*-test showed that the means of *h* and $\pi$ significantly differed between CSB (t = 32.678, df = 116.3, *p*-value < 0.01 for *h* value and t = 18.607, df = 126, *p*-value < 0.01 for $\pi$ value) and TAS (t = −29.507, df = 125.8, *p*-value < 0.01 for *h* value and t = −9.779, df = 135.6, *p*-value < 0.01 for $\pi$ value). Complex haplotype networks of both CR1 and CR2 were constructed from a large number of polymorphic sites and haplotypes, showing a striking star-shaped topology (Figure 2). The average sequence divergences between CR1 and CR2 (the paralogous CRs) of the same species (*p*-distance) were $0.39\% \pm 0.11\%$ and $0.37\% \pm 0.10\%$, respectively, whereas those of orthologous CRs in different species were $10.83\% \pm 0.58\%$ for CR1 and $17.08\% \pm 0.45\%$ for CR2 (Table 2).
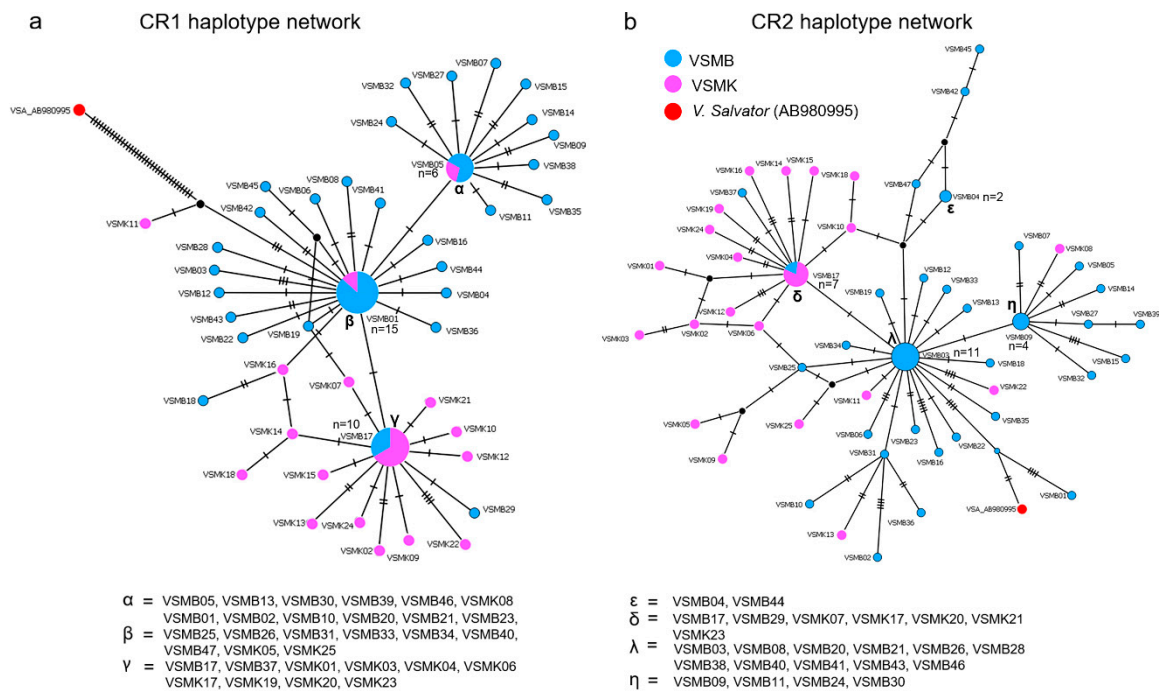
**Figure 2.** Haplotype network based on mitochondrial control region (mtCR) region sequence data of water monitors from Bang Kachao Peninsula (VSMB) and Varanus Farm Kamphaeng Saen (VSMK) populations, constructed using statistical parsimony with the TCS network. The numbers of individuals possessing haplotypes are indicated by different colors inside the circles. Missing haplotypes are indicated by black circles. (**a**) mtCR1 haplotype network (**b**) mtCR2 haplotype network.

**Table 2.** Percentage of D-loop sequence diversity for Asian water monitor (*Varanus salvator macromaculatus*; Deraniyagala 1944).

| Region | Within Group | Between Group |
|---|---|---|
| | *p*-Distance | *p*-Distance |
| CR1 | 0.39 ± 0.11 | - |
| CR2 | 0.37 ± 0.10 | - |
| CSB_CR1 | 0.40 ± 0.18 | - |
| CSB_CR2 | 0.15 ± 0.09 | - |
| TAS_CR1 | 011 ± 0.11 | - |
| TAS_CR2 | 0.21 ± 0.08 | - |
| CR1 * | 10.83 ± 0.58 | 25.11 ± 5.50 |
| CR2 * | 17.08 ± 0.45 | 38.88 ± 1.54 |
| CSB_CR1 * | 11.23 ± 1.96 | 34.70 ± 6.40 |
| CSB_CR2 * | 14.33 ± 1.88 | 32.49 ± 6.00 |
| TAS_CR1 * | 13.78 ± 0.17 | 41.10 ± 5.50 |
| TAS_CR2 * | 17.08 ± 0.45 | 38.88 ± 1.55 |

* Comparison between CRs of *Varanus salvator macromaculatus* with CRs of *V. salvator*, *V. salvator macromaculatus*, *V. salvator komaini*, *V. exanthematicus*, *V. komodoensis*, and *V. niloticus*.

Substitution saturation was estimated for both CR1 and CR2 datasets. No saturation was detected in CR2, as reflected by the linear correlation of the number of transitions and transversions plotted against sequence divergence (Figure 3) as well as from a significantly lower value of $I_{ss}$ as compared to $I_{ss.c}$ (Table 3). By contrast, in CR1, the number of transitions was higher than that of transversions, and substitution saturation occurred when the frequency of transitions exceeded the frequency of transversions (Figure 3).
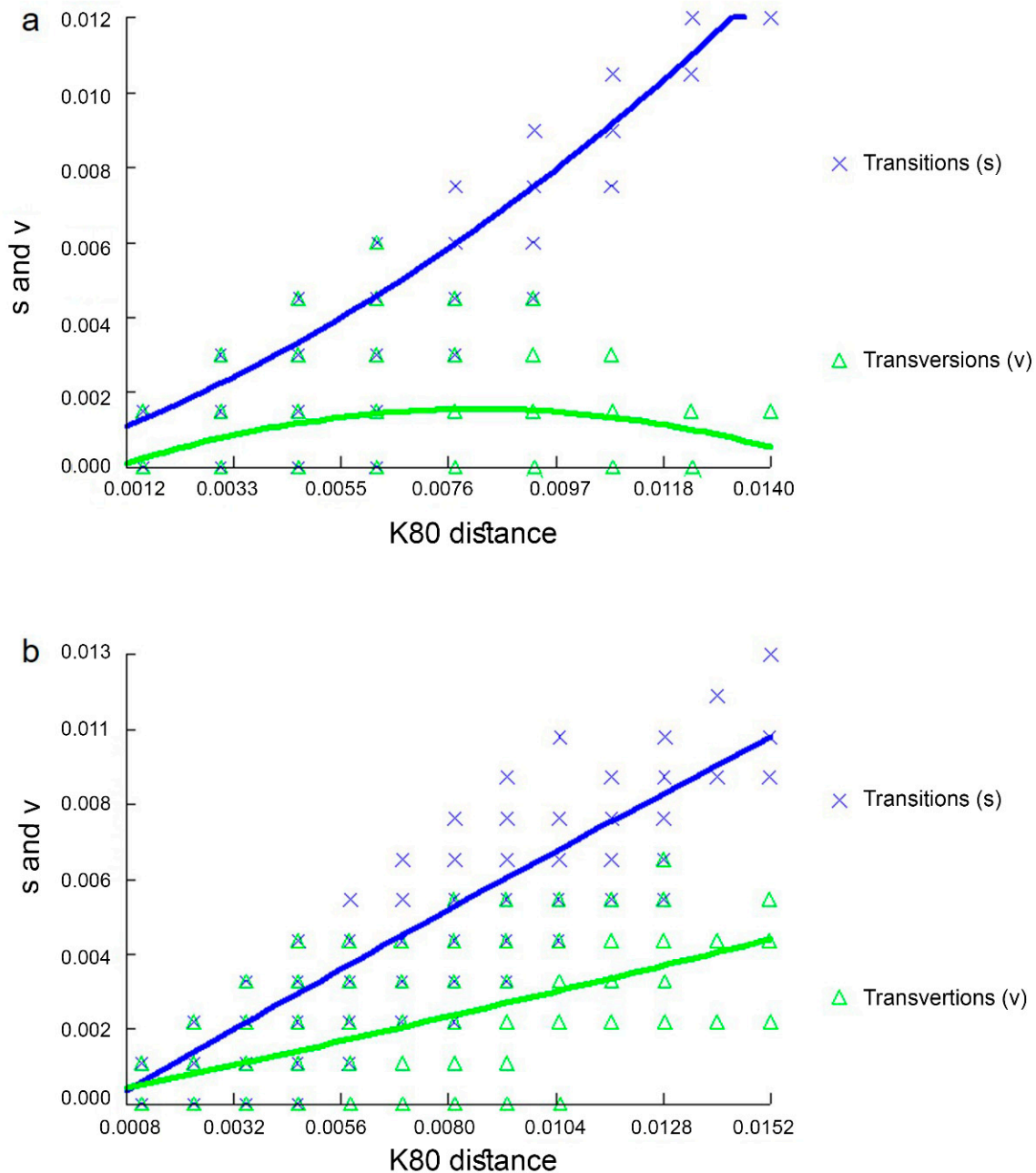
**Figure 3.** DAMBE7 substitution saturation plots for (**a**) CR1 and (**b**) CR2. Numbers of transitions (s) and transversions (v) are plotted against the K80 distance; lines indicate mean values (thick lines) and standard deviations (fine lines) of s and v.

*3.3. Phylogenetic Relationship Based on the Control Regions of V. salvator macromaculatus*

Phylogenetic analyses based on CR1 and CR2 were supported with high posterior probabilities and bootstrap values. Although the phylogenetic trees shared similar topologies, general differences existed (Supplementary Figures S1–S4). When combining both CR sequence datasets, paralogous CRs of an individual in the same dataset did not group together, while orthologous CRs of different individuals always clustered. When reconstructed separately using the separated regions (Supplementary Figures S5–S10), phylogenetic trees of TAS and CSB sequences shared the same topologies, indicating that orthologous copies from different individuals always clustered together rather than with paralogous copies from the same individuals. However, CR1 and CR2 from the same species are always clustered together (Supplementary Figures S1–S4).

**Table 3.** Substitution saturation analysis of CRs based on the index of substitution saturation as implemented in DAMBE7 [32].

| Region | Number of OUT [a] | $I_{ss}$ [b] | $I_{ss.cSym}$ [c] | Df [d] | *p*-Value [e] | $I_{ss.cAsym}$ [f] | df | *p*-Value |
|--------|-----------|------|----------|-----|----------|-----------|-----|----------|
| CR1 | 4 | 0.028 | 0.805 | 662 | <0.00001 | 0.774 | 662 | 0.0000 |
| | 8 | 0.033 | 0.766 | 662 | <0.00001 | 0.656 | 662 | 0.0000 |
| | 16 | 0.039 | 0.745 | 662 | <0.00001 | 0.535 | 662 | 0.0000 |
| | 32 | 0.046 | 0.719 | 662 | <0.00001 | 0.393 | 662 | 0.0000 |
| CR2 | 4 | 0.007 | 0.815 | 866 | <0.00001 | 0.784 | 866 | 0.0000 |
| | 8 | 0.008 | 0.780 | 866 | <0.00001 | 0.673 | 866 | 0.0000 |
| | 16 | 0.008 | 0.763 | 866 | <0.00001 | 0.560 | 866 | 0.0000 |
| | 32 | 0.009 | 0.738 | 866 | <0.00001 | 0.424 | 866 | 0.0000 |

[a] number of sequences used in random resampling; OTP: operational taxonomic unit, [b] index of substitution saturation, [c] critical value for a symmetrical tree topology, [d] degrees of freedom, [e] probability that $I_{ss}$ is significantly different from the critical value ($I_{ss.cSym}/I_{ss.cAsym}$), [f] critical value for an asymmetrical tree topology.

*3.4. Recombination Events in CRs*

To explain the presence of discordant signals between phylogenetic trees derived from CR1 and CR2 constructs, multiple recombination points were investigated using RDP software (Figure 4). Non-significant recombination events in CRs were observed at positions 138 bp and 672 bp in CR2 (MaxChi, *p* = 0.137; RDP, Geneconv, and Chimaera, no evidence of recombination was found) whereas no events were found in CR1. However, when we combined CR1 and CR2 of all individuals, a recombination event was observed at position 555 (Geneconv, *p* < 0.05) while no evidence of recombination was found in RDP, MaxChi, and Chimaera.
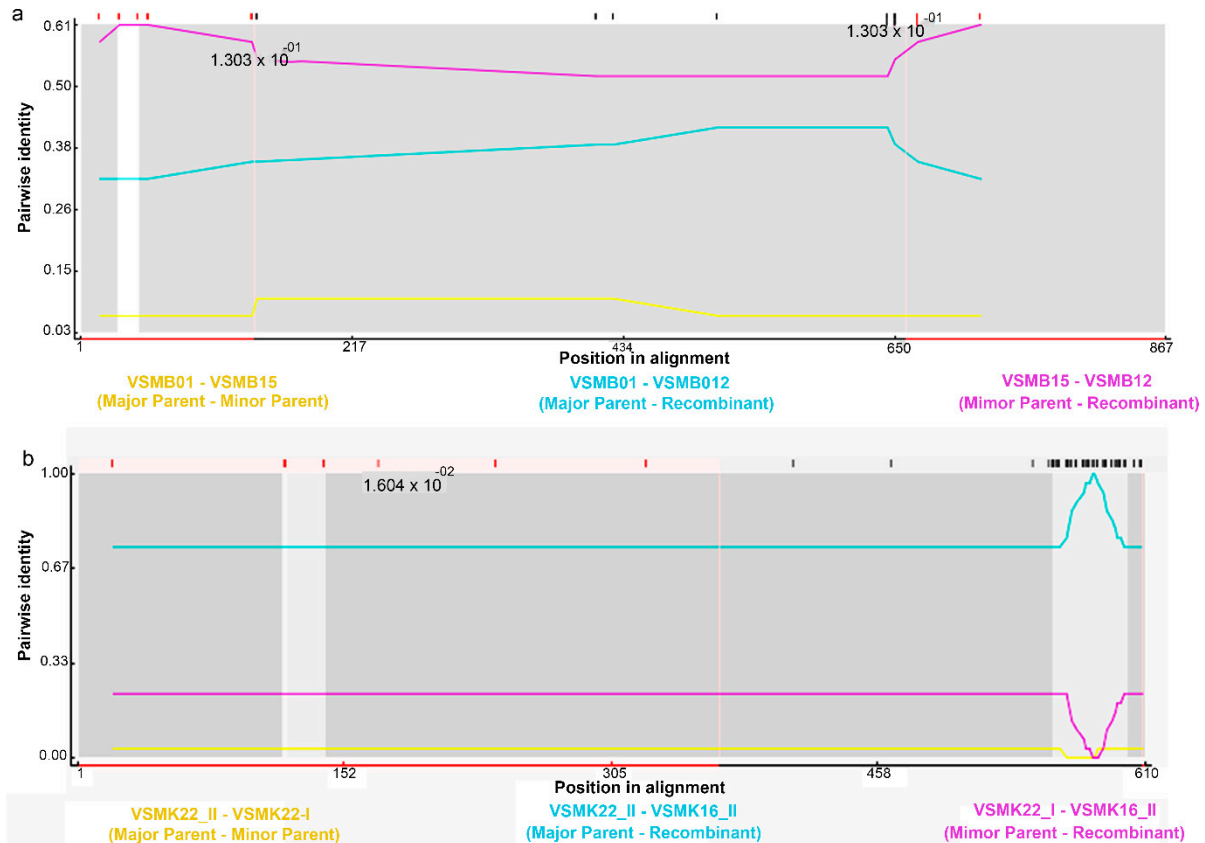


**Figure 4.** Assessment of recombination in duplicate CRs of all individuals: (**a**) CR2; (**b**) CRs detected using RDP software.

## 4. Discussion

Vertebrate mitogenome sequences are important systems that are predominantly utilized for molecular evolutionary studies, phylogenetics, and systematic taxonomy [3,5,7,44–50]. A special phenomenon has been observed in mitogenomes involving several vertebrates with duplicate CRs [9]. Comparisons of the four mitogenomes of varanids have revealed that all species possess duplicate CRs that tally with the process of mitogenomic rearrangement, and they can be reshuffled by investigating the use of PCR-based DNA marker analysis across 11 varanids [15]. The CR structures in vertebrate mitogenomes predominantly contain conserved sequences known as TASs and CSBs, which were observed in both CR1 and CR2. These conserved sequences are known to play important roles in the replication and expression of mitogenomic genes [6,51]. However, no CD was observed for either CR1 or CR2. This result suggests the plasticity of mitogenomes with the CD motif across vertebrates.

Entire sequences of CRs within species were highly similar, and the paralogs of CR1 and CR2 from each species showed a closer resemblance than those of their orthologs from other lineages. This suggests that CRs concertedly evolved in each species [9]. Duplicated CRs might remain conserved during reproduction and thus maintained in the mitogenome during speciation of varanids [6]. Alternatively, duplicate CRs probably play different roles in the replication of mitogenomes under evolutionary selective pressure and may have evolved independently within a particular species [52]. We observed this appearance in sequence divergences between CR1 and CR2 in 72 *V. salvator macromaculatus* individuals from the two populations [5]. The observed values were 5.91% and 17.4% between the two CRs. Homologs of CRs from different individuals were genetically more similar than paralogous CRs from the same individual within a species (Supplementary Figure S3), thereby agreeing with phylogenetic analyses, although both populations studied exhibited a high degree of population-level genetic diversity [5]. This suggests the independent evolution of evolutionary patterns for duplicate CRs within *V. salvator macromaculatus*. The CRs would have evolved independently within each varanid species after the divergence of species as a result of different mutations between CR1 and CR2 during evolution.

However, with evolutionary variations in characteristics, we found evidence of substantial variation in the inferred usage of CRs in *V. salvator macromaculatus*. Saturation analyses also showed that plots of transitions and transversions related linearly, with sequence divergence indicating no saturation in the CR2 data set [53]. The non-linear pattern in the CR1 data set suggested substitution saturation (Figure 3); therefore, CR2 may be a more informative phylogenetic marker at the population level in varanids. Different mitochondrial genome coding regions with diverse mutation rates and evolutionary trajectories are required to further elucidate the varanid phylogenetic lineage. Further analysis on their differential expression should be performed across varanid species to clearly understand the role of duplicate CRs. Although phylogenetic trees of CR1 and CR2 shared generally similar topologies, only minor differences existed in the placement of specimens (Supplementary Figures S1–S10). Additionally, ML and BI trees were constructed based on the TAS and CSB datasets derived from all individuals with three other variants. Phylogenetic trees of TAS and CSB sequences were largely congruent with each other, and homologs within an individual always formed closer clusters than their paralogs from other individuals. The phylogenetic analysis detected certain discordant signals among different CRs, suggesting that recombination might have reshaped the evolution of duplicate CRs in the mitogenomes of water monitors. The results of this study provide several recombination points for CR2. The breakpoints of recombination tend to occur at 138 bp and 672 bp in the TAS motif. It is essential to further investigate the recombination sites of CRs to better understand the evolutionary conflict and accurately detect the phylogenetic patterns [6,54,55]. This can be achieved by analyzing heterologous sequences that contribute to mitogenomic recombination [6]. In the mtDNA of water monitors, heterogeneous regions (A and T arrays) were detected in the VNTR motif [56]. There are four types of compound CT arrays. Different studies have identified the mitogenomic recombination

sites associated with heterologous in various vertebrates [6,56,57]. Specifically, VNTRs in CR2 of different species or individuals were heterologous with respect to their sequence size and motifs. This suggests that VNTRs have recombination roles. Further analyses with more varanid species are required to investigate more basic questions such as how mitochondrial genomes with duplicate CRs evolved and how the nucleotide sequences of duplicate CRs remained identical or highly similar over evolution under the concept of concerted and independent evolution.

## 5. Conclusions

This paper reports foundational knowledge on the dynamics of duplicated CRs in varanids mitogenomes. Our data suggest that these sequences might follow independent evolution within the same species. CRs seem to have acquired concerted evolution across different species. This hypothesis provides a baseline to study mitogenomic evolutionary events such as recombination, gene rearrangement, and concerted evolution between duplicates. A thorough understanding of nucleotide substitution in varanid CRs is important for advances in evolutionary model construction, with accurate evolutionary inferences to provide basic insights into the biology of mitogenomes.

## References

1. Taanman, J.W. The mitochondrial genome: Structure, transcription, translation and replication. *Biochim. Biophys. Acta Bioenerg.* **1999**, *1410*, 103–123. [CrossRef]
2. Gadaleta, M.C.; Noguchi, E. Regulation of DNA replication through natural impediments in the eukaryotic genome. *Genes* **2017**, *8*, 98. [CrossRef]
3. Areesirisuk, P.; Muangmai, N.; Kunya, K.; Singchat, W.; Sillapaprayoon, S.; Lapbenjakul, S.; Thapana, W.; Kantachumpoo, A.; Baicharoen, S.; Rerkamnuaychoke, B.; et al. Characterization of five complete *Cyrtodactylus* mitogenome structures reveals low structural diversity and conservation of repeated sequences in the lineage. *PeerJ* **2018**, *6*, e6121. [CrossRef] [PubMed]
4. Vinson, C.C.; Mangaravite, E.; Sebbenn, A.M.; Lander, T.A. Using molecular markers to investigate genetic diversity, mating system and gene flow of Neotropical trees. *Rev. Bras. Bot.* **2018**, *41*, 481–496. [CrossRef]
5. Wongtienchai, P.; Lapbenjakul, S.; Jangtarwan, K.; Areesirisuk, P.; Mahaprom, R.; Subpayakom, N.; Singchat, S.; Sillapaprayoon, S.; Muangmai, N.; Songchan, R.; et al. Genetic management of a water monitor lizard (*Varanus salvator macromaculatus*) population at Bang Kachao Peninsula as a consequence of urbanization with Varanus Farm Kamphaeng Saen as the first captive research establishment. *J. Zool. Syst. Evol. Res.* **2021**, *59*, 484–497. [CrossRef]
6. Zheng, C.; Nie, L.; Wang, J.; Zhou, H.; Hou, H.; Wang, H.; Liu, J. Recombination and evolution of duplicate control regions in the mitochondrial genome of the Asian big-headed turtle, *Platysternon megacephalum*. *PLoS ONE* **2013**, *8*, e82854. [CrossRef] [PubMed]
7. Singchat, W.; Areesirisuk, P.; Sillapaprayoon, S.; Muangmai, N.; Baicharoen, S.; Suntrarachun, S.; Chanhome, L.; Peyachoknagul, S.; Srikulnath, K. Complete mitochondrial genome of Siamese cobra (*Naja kaouthia*) determined using next-generation sequencing. *Mitochondrial DNA Part B Resour.* **2019**, *4*, 577–578. [CrossRef]
8. Lallemand, T.; Leduc, M.; Landès, C.; Rizzon, C.; Lerat, E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes* **2020**, *11*, 1046. [CrossRef]
9. Li, D.H.; Shi, W.; Munroe, T.A.; Gong, L.; Kong, X.Y. Concerted evolution of duplicate control regions in the mitochondria of species of the flatfish family Bothidae (Teleostei: Pleuronectiformes). *PLoS ONE* **2015**, *10*, e0134580. [CrossRef] [PubMed]
10. Uetz, P.; Freed, P.; Aguilar, R.; Hošek, J. The Reptile Database. 2021. Available online: http://www.reptile-database.org (accessed on 23 May 2006).

11. Ouwens, P.A. On a large *Varanus* species from the island of Komodo. *Bull. Jard. Bot. Buitenzorg.* **1912**, *6*, 1–3.

12. Kumazawa, Y.; Endo, H. Mitochondrial genome of the Komodo dragon: Efficient sequencing method with reptile-oriented primers and novel gene rearrangements. *DNA Res.* **2004**, *11*, 115–125. [CrossRef]

13. Fitzinger, L. *Neue Classification der Reptilien nach ihren natürlichen Verwandtschaften nebst einer Verwandschafts-Tafel und einem Verzeichnisse der Reptilien-Sammlung des K. K. Zoologischen Museums zu Wien*; J.G., Heubner: Wien, Austria, 1826; p. 66.

14. Kumazawa, Y. Mitochondrial genomes from major lizard families suggest their phylogenetic relationships and ancient radiations. *Gene* **2007**, *388*, 19–26. [CrossRef] [PubMed]

15. Amer, S.A.; Kumazawa, Y. Timing of a mtDNA gene rearrangement and intercontinental dispersal of varanid lizards. *Genes Genet. Syst.* **2008**, *83*, 275–280. [CrossRef]

16. Deraniyagala, P.E.P. Four New Races of the "Kabaragoya" Lizard *Varanus salvator*. *Spolia Zeylan.* **1944**, *24*, 59–62.

17. Hague, M.T.J.; Routman, E.J. Does population size affect genetic diversity? A test with sympatric lizard species. *J. Hered.* **2016**, *116*, 92–98. [CrossRef]

18. Supikamolseni, A.; Ngaoburanawit, N.; Sumontha, M.; Chanhome, L.; Suntrarachun, S.; Peyachoknagul, S.; Srikulnath, K. Molecular barcoding of venomous snakes and species-specific multiplex PCR assay to identify snake groups for which antivenom is available in Thailand. *Genet. Mol. Res.* **2015**, *14*, 13981–13997. [CrossRef] [PubMed]

19. Parham, J.F.; Feldman, C.R.; Boore, J.L. The complete mitochondrial genome of the enigmatic big-headed turtle (Platysternon): Description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. *BMC Evol. Biol.* **2006**, *6*, 11. [CrossRef]

20. Peng, Q.L.; Nie, L.W.; Pu, Y.G. Complete mitochondrial genome of Chinese big-headed turtle, *Platysternon megacephalum*, with a novel gene organization in vertebrate mtDNA. *Gene* **2006**, *380*, 14–20. [CrossRef] [PubMed]

21. Easterling, K.A.; Pitra, N.J.; Morcol, T.B.; Aquino, J.R.; Lopes, L.G.; Bussey, K.C.; Matthews, P.D.; Bass, H.W. Identification of tandem repeat families from long-read sequences of *Humulus lupulus*. *PLoS ONE* **2020**, *15*, e0233971. [CrossRef]

22. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]

23. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987.

24. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]

25. Clement, M.; Snell, Q.; Walker, P.; Posada, D.; Crandall, K. TCS: Estimating gene genealogies. 16th International Parallel and Distributed Processing Symposium. *Int. Proc.* **2002**, *2*, 184.

26. Dalgaard, P. Power and the computation of sample size. In *Introductory Statistics with R*; Springer: New York, NY, USA, 2008; pp. 155–162.

27. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

28. Salemi, M. Nucleotide substitution models. PRACTICE: The PHYLIP and TREEPUZZLE software packages. In *The Phylogenetic Handbook a Practical Approach to Phylogenetic Analysis and Hypothesis Testing*; Salemi, M., Vandamme, A.M., Eds.; Cambridge University Press: Cambridge, UK, 2003; pp. 88–97.

29. Felsenstein, J. Distance methods for inferring phylogenies: A justification. *Evolution* **1984**, *38*, 16–24. [CrossRef]

30. Xia, X.; Xie, Z.; Salemi, M.; Chen, L.; Wang, Y. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **2003**, *26*, 1–7. [CrossRef]

31. Xia, X.; Lemey, P. *Assessing Substitution Saturation with DAMBE. The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*; Cambridge University Press: Cambridge, UK, 2009; pp. 615–630.

32. Xia, X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **2018**, *35*, 1550–1552. [CrossRef]

33. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]

34. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Sys. Biol.* **2018**, *67*, 901. [CrossRef] [PubMed]

35. Tanabe, A.S. Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* **2011**, *11*, 914–921. [CrossRef] [PubMed]

36. Castoe, T.A.; Jiang, Z.J.; Gu, W.; Wang, Z.O.; Pollock, D.D. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE* **2008**, *3*, e2201. [CrossRef]

37. Chaiprasertsri, N.; Uno, Y.; Peyachoknagul, S.; Prakhongcheep, O.; Baicharoen, S.; Charernsuk, S.; Nishida, C.; Mutsuda, Y.; Koga, A.; Srikulnath, K. Highly species-specific centromeric repetitive DNA sequences in lizards: Molecular cytogenetic characterization of a novel family of satellite DNA sequences isolated from the water monitor lizard (*Varanus salvator macromaculatus*, Platynota). *J. Hered.* **2013**, *104*, 798–806. [CrossRef] [PubMed]

38. Martin, D.; Rybicki, E. RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* **2000**, *16*, 562–563. [CrossRef]

39. Smith, J.M. Analysing the mosaic structure of genes. *J. Mol. Evol.* **1992**, *35*, 126–129.

40. Sawyer, S. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **1989**, *6*, 526–538. [PubMed]

41. Posada, D.; Crandall, K.A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13757–13762. [CrossRef]

42. Martin, D.P.; Varsani, A.; Roumagnac, P.; Botha, G.; Maslamoney, S.; Schwab, T.; Kelz, Z.; Kumar, V.; Murrell, B. RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **2021**, *7*, veaa087. [CrossRef] [PubMed]

43. Ujvari, B.; Dowton, M.; Madsen, T. Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biol. Lett.* **2007**, *3*, 189–192. [CrossRef]

44. Srikulnath, K.; Thongpan, A.; Suputtitada, S.; Apisitwanich, S. New haplotype of the complete mitochondrial genome of *Crocodylus siamensis* and its species-specific DNA markers: Distinguishing *C. siamensis* from *C. porosus* in Thailand. *Mol. Biol. Rep.* **2012**, *39*, 4709–4717. [CrossRef] [PubMed]

45. Laopichienpong, N.; Muangmai, N.; Supikamolseni, A.; Twilprawat, P.; Chanhome, L.; Suntrarachun, S.; Peyachoknagul, S.; Srikulnath, K. Assessment of snake DNA barcodes based on mitochondrial *COI* and *Cytb* genes revealed multiple putative cryptic species in Thailand. *Gene* **2016**, *594*, 238–247. [CrossRef] [PubMed]

46. Laopichienpong, N.; Ahmad, S.F.; Singchat, W.; Suntronpong, A.; Pongsanarm, T.; Jangtarwan, K.; Bulan, J.; Pansrikaew, T.; Panthum, T.; Ariyaraphong, N.; et al. Complete mitochondrial genome of Mekong fighting fish, *Betta smaragdina* (Teleostei: Osphronemidae). *Mitochondrial DNA Part B Resour.* **2021**, *6*, 776–778. [CrossRef] [PubMed]

47. Lapbenjakul, S.; Thapana, W.; Twilprawat, P.; Muangmai, N.; Kanchanaketu, T.; Temsiripong, Y.; Unajak, S.; Peyachoknagul, S.; Srikulnath, K. High genetic diversity and demographic history of captive Siamese and Saltwater crocodiles suggest the first step toward the establishment of a breeding and reintroduction program in Thailand. *PLoS ONE* **2017**, *12*, e0184526. [CrossRef] [PubMed]

48. Ahmad, S.F.; Laopichienpong, N.; Singchat, W.; Suntronpong, A.; Pongsanarm, T.; Panthum, T.; Ariyaraphong, N.; Bulan, J.; Pansrikaew, T.; Jangtarwan, K.; et al. Next-generation sequencing yields complete mitochondrial genome assembly of peaceful betta fish, *Betta imbellis* (Teleostei: Osphronemidae). *Mitochondrial DNA Part B Resour.* **2020**, *5*, 3856–3858. [CrossRef] [PubMed]

49. Ariyaraphong, N.; Laopichienpong, N.; Singchat, W.; Panthum, T.; Ahmad, S.F.; Jattawa, D.; Duengkae, P.; Muangmai, N.; Suwanasopee, T.; Koonawootrittriron, S.; et al. High-level gene flow restricts genetic differentiation in dairy cattle populations in Thailand: Insights from large-scale mt D-Loop sequencing. *Animals* **2021**, *11*, 1680. [CrossRef]

50. Ariyaraphong, N.; Pansrikaew, T.; Jangtarwan, K.; Thintip, J.; Singchat, W.; Laopichienpong, N.; Pongsanarm, T.; Panthum, T.; Suntronpong, A.; Ahmad, S.F.; et al. Introduction of wild Chinese gorals into a captive population requires careful genetic breeding plan monitoring for successful long-term conservation. *Glob. Ecol. Conserv.* **2021**, *28*, e01675. [CrossRef]

51. Fonseca, M.M.; Harris, D.J.; Posada, D. The inversion of the control region in three mitogenomes provides further evidence for an asymmetric model of vertebrate mtDNA replication. *PLoS ONE* **2014**, *9*, e106654.

52. Eberhard, J.R.; Wright, T.F.; Bermingham, E. Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. *Mol. Biol. Evol.* **2001**, *18*, 1330–1342. [CrossRef]

53. Bronstein, O.; Kroh, A.; Haring, E. Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. *BMC Evol. Biol.* **2018**, *18*, 80. [CrossRef]

54. Abbott, C.L.; Double, M.C.; Trueman, J.W.; Robinson, A.; Cockburn, A. An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in *Thalassarche albatrosses*. *Mol. Ecol.* **2005**, *14*, 3605–3613. [CrossRef]

55. Morris-Pocock, J.A.; Taylor, S.A.; Birt, T.P.; Friesen, V.L. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Evol. Biol.* **2010**, *10*, 14. [CrossRef]

56. Hoarau, G.; Holla, S.; Lescasse, R.; Stam, W.T.; Olsen, J.L. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Mol. Biol. Evol.* **2002**, *19*, 2261–2264. [CrossRef]

57. Tsaousis, A.D.; Martin, D.P.; Ladoukakis, E.D.; Posada, D.; Zouros, E. Widespread recombination in published animal mtDNA sequences. *Mol. Biol. Evol.* **2005**, *22*, 925–933. [CrossRef] [PubMed]