Published in final edited form as:

Mol Phylogenet Evol. 2009 December ; 53(3): 784–791. doi:10.1016/j.ympev.2009.07.037.

Evolution of serum albumin intron-1 is shaped by a 5' truncated non-long terminal repeat retrotransposon in western Palearctic water frogs (Neobatrachia)

Jörg Plötner^{1,*}, Frank Köhler^{1,2}, Thomas Uzzell³, Peter Beerli⁴, Robert Schreiber¹, Gaston-Denis Guex⁵, and Hansjürg Hotz^{1,5}

¹ Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstraße 43, 10115 Berlin, Germany ² Department of Research, Australian Museum, 6 College Street, Sydney 2010 NSW, Australia ³ Academy of Natural Sciences, Laboratory for Molecular Systematics and Ecology, 1900 B. F. Parkway, Philadelphia 19103, USA ⁴ School of Computational Science and Department of Biological Science, Florida State University, Tallahassee, Florida 32306-4120, USA ⁵ Universität Zürich-Irchel, Zoologisches Institut, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Abstract

A 5' truncated non-LTR CR1-like retrotransposon, named *RanaCR1*, was identified in the serum albumin intron-1 (SAI-1) of at least seven species of western Palearctic water frogs (WPWF). Based on sequence similarity of the carboxy-terminal region (CTR) of ORF2 and the highly conserved 3' untranslated region (3' UTR), *RanaCR1*-like elements occur also in the genome of *Xenopus tropicalis* and *Rana temporaria*. Unlike other CR1 elements, *RanaCR1* contains a CA microsatellite in its 3' UTR. The low nucleotide diversity of the 3' UTR compared to the CTR and to SAI-1 suggests that this region still plays a role in WPWF, either as a structure-stabilizing element, or within a species-specific transcriptional network. Length variation of water frog SAI-1 sequences is caused by deletions that extend in some cases beyond the 5' or 3' ends of *RanaCR1*, probably a result of selection for structural and functional stability of the primary transcript. The impact of *RanaCR1* on SAI-1 evolution is also indicated by the significant negative correlation between the length of both SAI-1 and *RanaCR1* and the percentage GC content of *RanaCR1*. Both SAI-1 and *RanaCR1* sequences support the sister group relationship of *R. perezi* and *R. saharica*, which are placed in the phylogenetic tree at a basal position, the sister clade to other water frog taxa. It also supports the monophyly of the *R. lessonae* group; of Anatolian water frogs (*R. cf. bedriagae*), which are not conspecific with *R. bedriagae*; and of the European *ridibunda* group. Within the *ridibunda* clade, Greek frogs are clearly separated, supporting the hypothesis that Balkan water frogs represent a distinct species. Frogs from Atyrau (Kazakhstan), the type locality of *R. ridibunda*, were heterozygous for a *ridibunda* and a *cf. bedriagae* specific allele.

Keywords

serum albumin intron-1; CR1-like retrotransposon; microsatellite; western Palearctic water frogs; *Rana* (*Pelophylax*); Neobatrachia

*Corresponding author: Telephone: 0049 30 20938508, Fax: 0049 30 20938528, joerg.ploetner@mfn-berlin.de.

Introduction

Spliceosomal introns are parts of all eukaryote genomes so far investigated. They are generally composed of quasi-random sequences and lack open reading frames (reviewed by Roy and Gilbert, 2006). Introns are excised ('spliced') out of RNA transcripts of coding genes prior to protein synthesis and thus are usually considered non-functional and selectively neutral, but conserved sequence motifs, especially around the intron-exon boundaries, indicate that there are functional constraints linked with the spliceosomal machinery. The formation of secondary structures in pre-mRNA may also be functional (reviewed by Buratti and Baralle, 2004).

The mechanisms of intron evolution are largely unknown. Comparative studies using intron sequences of closely and distantly related taxa may help to identify functional and nonfunctional structural motifs important for understanding these mechanisms. The western Palearctic water frog (WPWF) group includes at least seven closely related species (Plötner, 2005), most of which originated in the upper Miocene and middle Pliocene/early Pleistocene (Beerli et al., 1996; Plötner et al., in press). It thus provides an opportunity to study intron evolution on a 'fine scale', within the last 10 million years (My).

Here we present first results of such a comparative study of intron-1 of the serum albumin gene (SAI-1) of WPWFs. We also describe a non-long terminal repeat (non-LTR) retrotransposon that is embedded in the SAI-1 of the water frog species investigated. This non-LTR retrotransposon was classified as a chicken repeat (CR) 1-like long interspersed nuclear element (LINE). CR1 retrotransposons (Stumph et al., 1981) and related non-LTR retroelements are widely distributed in the genomes of vertebrates and invertebrates (e.g. Eickbush, 1994; Haas et al., 2001; Deininger and Batzer, 2002; Deininger et al., 2003; Shedlock, 2006; Kordiš et al., 2006; Novikova et al., 2007; Shedlock et al., 2007; and literature cited therein). They are increasingly used as characters for phylogenetic reconstructions (e.g., John et al., 2005; Kordiš et al., 2006; Shedlock, 2006; Watanabe et al. 2006; Kaiser et al., 2007; Shedlock et al., 2007; Treplin and Tiedemann, 2007). The primary advantage of retroelement insertions for such studies is the high likelihood that, if two genomes share a mobile element at the same locus, the mobile element and insertion are identical by descent (e.g., Deininger and Batzer, 2002).

We compared the structure and base composition of the SAI-1 and CR1-like sequences and discuss the impact that *Rana*CR1 has had on the evolution of SAI-1 in WPWF. We also used sequences of both SAI-1 and the inserted CR1-like retroelement to test hypotheses on water frog systematics proposed on the basis of protein electrophoretic data (Beerli et al., 1996) and mitochondrial sequences (Plötner, 1998; 2005; Plötner and Ohst, 2001; Plötner et al., 2001; 2007).

2. Materials and methods

2.1. Samples

We analysed SAI-1 and *Rana*CR1 sequences of 24 WPWFs from 21 localities (Table 1) and one eastern Palearctic water frog (a *R. nigromaculata* from North Korea, EMBL FN432386).

2.2. DNA isolation and primer development

Total genomic DNA was extracted from pieces of muscle tissue taken from ethanol preserved specimens by use of a DNA extraction kit (Qiagen, Hilden, Germany) following the standard protocol for preserved tissues. Primers for PCR amplification and sequencing were designed using the serum albumin cDNA of *Rana shqipERICA*. For this purpose, RNA was isolated from fresh liver of an adult male *R. shqipERICA* collected near Bushat, Albania (41°57'N, 19°32'E). Tissue was ground in liquid N₂ and transferred to a guanidinium isothiocyanate solution

(Chirgwin et al., 1979). This solution was centrifuged through a CsCl step gradient; the RNA pellet was recovered and polyadenylated RNA was isolated by use of oligo-(dT) cellulose (Stratagene). Reverse-transcribed cDNA was ligated first to adaptor arms and then inserted into phage Lambda Zap II following Stratagene protocols. Phages containing albumin cDNA inserts were isolated by plating a small number of phage particles on a lawn of XL1 Blue cells (Stratagene). Albumin-positive plaques were identified on nylon membranes by hybridization with an αP^{32} -labeled partial albumin cDNA from *Rana catesbeiana* (GenBank M38195; Averyhart-Fullard and Jaffe, 1990). Positive inserts were isolated in phagemid pBluescript by *in vivo* excision. The longest insert was sequenced in both directions using Sanger's dideoxy method (cf. Barnes et al., 1983). Vector primers SK and T₇ and five additional pairs of 15-mer or 16-mer primers were sufficient to sequence the entire insert. An area with bad compression artifacts was resequenced using dITPs.

The cDNA sequence of *R. shqiperica* (U40452) was aligned with the complete sequence of human serum albumin (M12523; Minghetti et al., 1986), paying particular attention to the intron-exon borders of the human albumin gene, after which it was possible to delimit the exons in the ranid cDNA sequence. Exon-primers (Ex) were designed with Primer3 (<http://primer3.sourceforge.net>; Rozen and Skaletsky, 2000). Forward primers are situated in exon 1, reverse primers in exon 2. Seven additional internal primers (Int) were designed after initial amplification and partial sequencing of albumin intron 1 (Table 2).

2.3. PCR, sequencing, and cloning

The complete intron-1 of the serum albumin gene was amplified and sequenced using the primers listed in Table 2. PCR amplifications were conducted in 25 μ l volumes containing 1x PCR buffer, 200 μ M each dNTP, 2.0 mM MgCl₂, 0.5 μ M each primer, 1.25 units of Taq polymerase (New England Biolabs, Ipswich, MA, USA), and approximately 50 ng of DNA. An initial denaturation step of 3 min at 96 °C was followed by 35 cycles of 60 s each at 94 °C, 50–55 °C, and 72 °C, and a final extension step of 5 min at 72 °C. PCR products were directly cycle sequenced using PCR primers and BigDye terminator chemistry on a 3130XL Genetic Analyzer (Applied Biosystems). PCR products that could not be directly sequenced because of heterozygosity were cloned. For this, the double stranded intron fragments were ligated into the pDrive cloning vector by use of the Qiagen PCR cloning kit following the standard protocol. Recombinant bacterial colonies were identified by blue-white screening. DNA of positive clones was amplified by *in vitro*-DNA replication (TempliPhi Amplification Kit, Amersham Biosciences, USA) and purified by use of a nucleo-spin DNA clean up kit (Macherey and Nagel, Düren, Germany). Sequencing of cloned fragments was performed as given above for PCR products using standard M13 primers contained by the cloning vector. Eight positive clones of each PCR product were sequenced in both directions in order to correct for polymerase-errors and to recover any polymorphisms present.

2.4. Sequence analysis and statistical procedures

All intron sequences were screened for interspersed repeats and low complexity DNA sequences using the program Repeat Masker (Institute for Systems Biology; <http://www.systemsbio.org>) with the options: search engine=cross match; speed/sensitivity=slow; DNA source=vertebrate.

Nucleotide and amino acid sequences were initially aligned using ClustalW (Higgins et al., 1994). Subsequently, the alignment (Appendix A, Supplementary data) was improved manually. Gaps were inserted to maximize the alignment identities. Uncorrected pairwise distances were calculated with MEGA 4.0 (Tamura et al., 2007); sites with gaps and missing data were completely deleted from the data set. DnaSP 5.0 (Librado and Rozas, 2009) was used to analyse DNA polymorphism; McDonald-Kreitman (MK) tests (McDonald and

Kreitman, 1991), conducted in DnaSP, were performed to test for selective neutrality of the carboxy-terminal region (CTR). Nucleotide diversity (π) was analysed at individual sites (column by column); sites with alignment gaps or missing data were excluded only in pairwise comparisons. A signed rank test, implemented in the software packet Statgraphics Plus 4.1 (Statistical Graphics Corp.; StatPoint Inc.; Herndon, Virginia, USA), was applied to test for differences in pairwise uncorrected substitution rates between the CTR and 3' UTR.

For sequence comparisons we used nucleotide and amino acid sequences of *Xenopus tropicalis* (e.g., DN044719, UPI00004, D7CA6), *Gallus gallus* (AAA49027.1; Burch et al., 1993), *Platemys spixii* (BAA88337.1; Kajikawa et al., 1997), and *Strongylocentrotus purpuratus* (XP_001194331.1; Towle and Smith, unpublished) available at the EMBL nucleotide sequence database. Plausible amino acid sequences for the CTR of these CR1 elements and *Rana*CR1 (pos. 318–710, Appendix A, Supplementary data), adjusting for frame-shift mutations, were aligned with Clustal W (MEGA) or MacClade 4.08. Genetic relationships of CR1 elements were analysed with Neighbor Joining (NJ) as implemented in the program MEGA 4.0 on the basis of the amino acid alignment using both uncorrected p and more sophisticated JTT distances (Jones et al., 1992). Tree robustness was evaluated by bootstrapping (Felsenstein, 1985) with 10^3 replicates.

To find appropriate models of sequence evolution for the different domains (SAI-1, *Rana*CR1, CTR) we used hierarchical likelihood ratio tests (hLRTs) and the Bayesian Information Criterion (BIC) implemented in the program Modeltest 3.8 (Posada and Crandall, 1998). Model parameters (site specific substitution rates, base frequencies, the shape parameter α of a gamma distribution for substitution rates, and the proportion of invariable sites) were estimated from the data. Signed rank test, regression and correlation analysis were performed with the software packet Statgraphics Plus 4.1. Phylogenetic relationships of WPWFs were analysed on the basis of the combined SAI-1 and *Rana*CR1 sequences with the Maximum Likelihood (ML) algorithm using the program PAUP* 4.0b10 (Swofford 2003). The mutation model chosen was HKY (Hasegawa et al., 1985) + G to allow for site variation. The mutation model parameters were evaluated using an NJ-tree and then fixed. The best phylogeny was found using the heuristic search under the ML criterion. Branch support was evaluated by bootstrapping (Felsenstein, 1985) with 10^3 replicates. For each replicate the mutation model was estimated and an ML search performed. The clade support was then expressed using a 50% majority-rule consensus tree.

3. Results

3.1. Structure of SAI-1 and *Rana*CR1

Intron 1 of the serum albumin gene in western Palearctic water frogs and the single eastern Palearctic water frog species (*R. nigromaculata*) examined has typical exon/intron splice site junction sequences (exon-1/GT for the donor site, AG/exon-2 for the acceptor site). An adenine (an essential element for splicing) is located in a pyrimidine rich region 13–14 nt upstream of the acceptor site. The BDGB program (http://www.fruitfly.org/seq_tools/splice.html) predicted several additional splice sites within the intron 1 sequence of several WPWF species. Analysis with the POLYADQ program (Tabaska and Zhang, 1999) identified one potential poly(A) signal (ATTAAA) at pos. 177–182 within the SAI-1 sequences of *R. saharica*, *R. epeirotica*, *R. bedriagae*, and *R. ridibunda* (Appendix A, Supplementary data).

The intron shows marked length polymorphism (Table 3) caused by a non-LTR retrotransposon that was identified by RepeatMasker as a member of the chicken repeat (CR) 1 family (Fig. 1). We therefore call this retroelement *Rana*CR1. *Rana*CR1 is present in SAI-1 of all the WPWFs but absent or else extremely reduced in SAI-1 of *R. nigromaculata*. The mean GC

content of *Rana*CR1 is about 20% higher than that of the surrounding intron sequence (pos. 1–291, pos. 796–1127) (Table 3).

The sequence motif AATAT at both flanks of *Rana*CR1 (pos. 287–291 and 796–800) is thought to represent the target site duplication (TSD), a typical result of the insertion event. At the 3' end of *Rana*CR1 the target site is followed by a 15 bp sequence that represents an inverted repeat; its counterpart is located upstream of the presumptive 5' end (Fig. 1).

*Rana*CR1, like most non-LTR retrotransposons, is truncated at the 5' end. It contains a domain homologous to the CTR of open reading frame 2 (ORF2) of complete CR1 elements. BLAST searches (Altschul et al., 1997) using the nucleotide and translated amino acid sequence of the CTR of *R. ridibunda* from Poland (17-PL) and *R. saharica* from Algeria (21-DZ) showed best matches with sequences obtained from the genomes of *X. tropicalis*, *G. gallus*, *P. spixii*, and *S. purpuratus* (for accession numbers see 2.4.). In many water frogs investigated, the ORF of the CTR is interrupted by frame shifts or in-frame stop codons (Appendix B, Supplementary data). NJ analyses based on the amino acid alignment of the CTR revealed a sister group relationship between the WPWFs and *X. tropicalis* under both p and JTT matrix based distances (Appendix B, Supplementary data) which emphasizes that *Rana*CR1 and the *Xenopus* elements both belong to the CR1 class of non-LTR retrotransposons. At the amino acid level the CTR of *Rana*CR1 shares about 60% sequence identity with homologous sequences of *X. tropicalis*.

After a stop codon (TAA) the CTR of most sequences is followed by a direct CA repeat that belongs to the 3' UTR (Fig. 1; Appendix A, Supplementary data). In sequences of the *R. lessonae* group the repeat is modified by an A>G transition at position 721. Like most CR1 elements *Rana*CR1 ends in a perfect octameric direct repeat (AACTATGT)₂, except those elements of Greek *R. ridibunda* (18-GR, 19-GR) and Italian water frogs (*R. lessonae* sensu lato) from Metaponto (11-IT) and Tarsia (12-IT). Greek *R. ridibunda* show a deletion that starts directly after the CA repeat and extends 63 nucleotides (pos. 796–865) downstream of the octameric repeat, whereas Italian frogs possess only one copy of the octameric sequence motif. Deletions that extend beyond the putative 5' target site were found in Italian and Central European *R. lessonae* (Appendix A, Supplementary data).

3.2. Patterns of sequence evolution

Nucleotide diversity (π) amounted to 0.033 in SAI-1 and 0.038 in *Rana*CR1. The π -values for the CTR and the highly conserved 3' UTR are 0.046 and 0.023, respectively. The results of the MK tests (data not shown) indicate no departure from a neutral model of sequence evolution in the CTR as expected for a region that is no longer translated. Interspecific sequence divergence (uncorrected p distances) for SAI-1 ranged between 0.3% (*R. cf. bedriagae* – allele *a* of *R. ridibunda* from the type locality) and 5.9% (*R. saharica*–*R. cf. bedriagae*); for *Rana*CR1, divergences ranged between 0.7 % (allele *b* of *R. ridibunda* from the type locality–*R. bedriagae*) and 7.8% (*R. cretensis*–*R. perezi* and *R. cretensis*–*R. saharica*). A signed rank test revealed significant differences in the uncorrected substitution rates obtained for the CTR and the 3'UTR of *Rana*CR1 ($p < 0.05$).

The best-fit model of sequence evolution for SAI-1 was the HKY model (Hasegawa et al. 1985). For *Rana*CR1 Modeltest proposed the similar HKY+G model with a gamma distributed shape parameter $G=0.32$. There is a statistically significant relationship at the 99% confidence level ($r=0.88$; ANOVA: F-ratio=373.9; $P < 0.001$) between the HKY distances (SAI-1) and HKY+G distances (*Rana*CR1) calculated for all pairwise sequence comparisons except the extremely truncated *R. lessonae* specific sequences (Appendix C, Supplementary data).

The impact of the transposition event on SAI-1 evolution is evidenced by a significant positive correlation between the size of *RanaCR1* and the size of SAI-1 ($r=0.95$; $p<0.001$). Moreover, the relative GC content of *RanaCR1* is significantly negatively correlated with the size of both *RanaCR1* ($r=-0.68$; $p<0.05$) and SAI-1 ($r=-0.77$; $p<0.001$).

3.3 Phylogenetic analysis

In the ML tree calculated on the basis of the combined SAI-1 and *RanaCR1* sequences, *R. saharica* and *R. perezi* branched off first as sister species in a clade that is strongly supported as the sister group of the remaining taxa (Fig. 2). The monophyly of *R. saharica*, *R. perezi*, the *R. lessonae* group, *R. cretensis*, Anatolian frogs (*R. cf. bedriagae*) including allele a of *R. ridibunda* (15-KZ) from the type locality (Atyrau, Kazakhstan), *R. epeirotica*, and European *R. ridibunda* together with the second nuclear haplotype of *R. ridibunda* from Atyrau (allele b) is recovered with strong bootstrap support ($>85\%$). There is also high bootstrap support (0.95) for a sister group relationship between *R. bedriagae* and the European *R. ridibunda* clade.

Within the European *R. ridibunda* clade, Greek individuals (18-GR, 19-GR) are clearly separated from Central European individuals (16-DE, 17-PL) with a bootstrap support of 0.87. The distinct status of Greek *R. ridibunda* is also expressed in the structural features of *RanaCR1*, especially in the deletion that extends from pos. 796 to 865 (Table 3; Appendix A, supplementary data). *R. epeirotica* represents the sister taxon of the *R. ridibunda*+*R. bedriagae* clade, although this branching pattern is only weakly supported. The phylogenetic positions of the *R. lessonae* group, *R. cretensis*, and the Anatolian group (*cf. bed, rid* 15-KZ) are not well resolved either.

4. Discussion

4.1 Structure of *RanaCR1* and its potential impact on SAI-1 evolution

RanaCR1 is the first non-LTR retroelement detected in the genome of neobatrachian anurans. Like most transposable elements (TEs), *RanaCR1* is 5' truncated probably resulting from abortive RNA reverse transcription, when the reverse transcriptase dissociates from its RNA template before having completed cDNA synthesis (e.g., Silva and Burch, 1989; Eickbush, 1994; Burch et al., 1993; Kajakawa et al., 1997; Poulter et al., 1999). As our data show, length variation of TEs may also be caused by deletions that occurred after the transposition event and extend beyond the 5' and 3' border of the retroelement. These deletions are thought to be linked to selection in the context of genomic instability caused by the insertion event (e.g., Kazazian and Goodier, 2002; Symer et al., 2002; Gilbert et al., 2002; 2005).

The insertion of retrotransposons into introns may lead to mis-splicing by introducing alternative splice sites, such as those in *RanaCR1*, that can result in improperly spliced transcripts (e.g., Deininger et al., 2003; Ostertag and Kazazian, 2005; Belancio et al., 2006), or by direct changes of structural motifs that act as recognition or binding sites for the spliceosomal machinery, or because of conformational changes in the pre-mRNA that may cause a decreasing or spurious transcription of the gene or a reduced stability of the primary transcript (Ostertag and Kazazian, 2005; Slotkin and Martienssen, 2007). Selection for structural and functional stability of serum albumin transcripts is also indicated by the significant negative correlation between the size of both SAI-1 and *RanaCR1* and the percentage GC content of *RanaCR1*. As an alternative explanation, deletions may be caused by unequal homologous recombination (Kazazian, 2004) because retroelements are recombination hot spots (e.g., Edelman et al., 1989; Burwinkel and Kilimann, 1998; Segal et al., 1999; Sen et al., 2006).

Retroelements are often associated with microsatellites (e.g. Beckmann and Weber, 1992; Nadir et al., 1996); it is hypothesized that A-rich microsatellites were generated by a 3' extension of retrotranscripts, similar to mRNA polyadenylation, or by nontemplated additions of nucleotides by reverse transcriptases (e.g., Luan and Eickbush, 1995; Kapitanov and Jurka, 2003). To our knowledge *RanaCR1* is the first case of a CR1 element in which (1) the 3' UTR contains both a microsatellite and an octameric repeat, and (2) the microsatellite is integrated upstream from the octameric repeat. Single microsatellites (e.g. BFG30 and Raja09), isolated from the genome of the brown frog species *Rana temporaria* (Matsuba and Merilä, 2008), are also integrated in the 3' UTR of retroelements similar to *RanaCR1* (Plötner, unpublished results). Another example of a microsatellite (CCTCT)_n located in the 3' UTR of LINEs comes from the Maui element of puffer fish (Poulter et al., 1999), which does not belong to the CR1 clade (Albalat et al., 2003). As these examples indicate, retrotransposition seems not to be a general mechanism that is responsible for microsatellite genesis associated with TEs.

All evidence indicates that the insertion of *RanaCR1* into SAI-1 was a singular event and that the sequences we have observed are all orthologous (all variants trace back to one and the same ancestral retroelement). Like most CR1 elements, *RanaCR1* is thought to be nonfunctional both because it is truncated at the 5' end and because of deletions and stop codons within its ORF that indicate lack of selective constraint. Assuming that *RanaCR1* is completely non-functional, it is difficult to explain why its CTR and 3' UTR evolve at obviously different rates within WPWFs. The high conservation of the 3' UTR of *RanaCR1* and its capacity to form hairpin structures (Plötner, unpublished data) suggests that this region probably still plays a role, as a structure-stabilizing element, as a recognition site for a protein involved in the regulation of chromatin structure as suggested by Stumph et al. (1983; 1984), or as a regulatory element in the context of RNA-mediated gene regulation. Recent analyses of human and plant genomic data revealed that TEs are an important source and target of small RNAs (Watanabe et al., 2006; Piriyaongsa and Jordan, 2007; 2008; reviewed by Slotkin and Martienssen, 2007). Whether the 3' UTR of *RanaCR1* really contributes to gene regulation at the transcriptional or post-transcriptional level as shown for other TEs (Watanabe et al., 2006; Piriyaongsa and Jordan, 2007; 2008; Piriyaongsa et al., 2007; reviewed by Zaratiegui et al. 2007; Feschotte, 2008) remains to be tested experimentally.

4.2. Implications for water frog systematics

Because *RanaCR1* was found in the SAI-1 of species spanning the deepest phylogenetic splits within WPWFs, it must have been present in their last common ancestor, i.e. in a form that lived more than 10 My ago (Uzzell, 1982; Beerli et al., 1996; Plötner et al., in press). It also seems possible, however, that the insertion event predates the split of eastern and western Palearctic water frogs: there is a possible imperfect TSD (AATTT___AATAT) and a putative 3 bp relict (pos. 793–795) of the octameric repeat present in the SAI-1 of the eastern Palearctic species *R. nigromaculata*. Investigation of SAI-1 sequences from other members of the eastern Palearctic water frog clade and from other clades in the Holarctic *Rana* radiation may help to resolve this question.

The results of phylogenetic analyses are extensively consistent with former hypotheses on water frog systematics based on immunological data (Uzzell, 1982), protein electrophoretic data (Beerli et al., 1996), and mitochondrial sequences (Plötner, 1998; 2005; Plötner and Ohst, 2001; Plötner et al. 2007; Lymberakis et al., 2007). In accordance with mitochondrial genealogies, all phylogenetic reconstructions with SAI-1 and *RanaCR1* sequences support the monophyly of *R. saharica*, *R. perezi*, *R. cretensis*, the *R. lessonae* group, *R. epeirotica*, and the *R. ridibunda* group sensu lato. ML and Bayesian trees, unlike MP trees based on mtDNA, place *R. epeirotica* closer to the *R. ridibunda*/*R. bedriagae* clade than to *R. cretensis* (Plötner et al., 2007). The well-supported basal position and sister group relationship of *R. saharica* and *R.*

perezi (Fig. 2) is consistent with mitochondrial phylogenies (e.g., Plötner, 2005; Plötner et al. 2007; Lymberakis et al., 2007). The relationships of most of the other taxa, however, are not well resolved. Short branches might be attributed to a rapid radiation that makes a resolution of phylogenetic relationships difficult.

Like in trees based on mtDNA data, *R. bedriagae* from Syria (01-SY) and Anatolian water frogs sensu lato (*R. cf. bedriagae*, 02-GR, 03-TR) are always placed in different clades. While Syrian *R. bedriagae* is closely related to European *R. ridibunda* (including allele *b* of the nuclear haplotype 15-KZ found at Atyrau, the type locality of *R. ridibunda*), Anatolian frogs are placed in a distinct clade together with allele *a* of haplotype 15-KZ. Thus, the SAI-1 and *RanaCR1* sequence data support the hypothesis that Anatolian frogs are not conspecific with *R. bedriagae* from the Near East (Plötner et al., 2001, in press).

All four Atyrau frogs investigated were heterozygous, possessing two SAI-1 alleles: one characteristic of Anatolian frogs, the other closely related to alleles found in individuals of the Central European *R. ridibunda* stock. Although the sample size is limited, this finding together with mitochondrial data (Plötner and Litvinchuk, unpubl.) argues for the existence of a hybrid zone east of the Caspian Sea, similar to hybrid zones found in eastern Greece (Hotz et al., unpublished data) and Italy (Santucci et al., 1996).

In summary, our results demonstrate the utility of both SAI-1 and *RanaCR1* sequences for genomic, evolutionary and systematic studies in water frogs. Further investigations on the albumin gene and the detected retroelement will certainly provide new insights in the evolutionary history and genomic structure not only of this interesting anuran group but also of other anuran taxa.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

For providing water frog samples, we thank Rainer Günther (Berlin) and Dirk Schmeller (Saint Giron). Tissue samples from Greece were made available by the Greek Ministry of Rural Development and Food, kindly mediated by Doris Tippmann (Embassy of the FRG, Athens). We are grateful to Nils Hof (Berlin) for technical assistance. Two anonymous reviewers provided constructive criticism. This work was supported by the Deutsche Forschungsgemeinschaft (grants PL 213/3-1, 3-2, 3-3) and the Swiss National Fund (grants 31-37579.93, 31-59144.99, 31-103903/1, and 31-64004.00). Peter Beerli was partly supported by the joint NSF/NIGMS Mathematical Biology program under NIH grant R01 GM 078985.

References

- Albalat R, Permany J, Cañestro C, Martínez-Mir A, González-Angulo O, González-Duarte R. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. *Cell Mol Life Sci* 2003;60:803–809. [PubMed: 12785727]
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
- Averyhart-Fullard V, Jaffe RC. Cloning and thyroid hormone regulation of albumin mRNA in *Rana catesbeiana* tadpole liver. *Mol Endo* 1990;4:1556–1563.
- Barnes, WM.; Bevan, M.; Son, PH. Kilo-sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. In: Wu, RL.; Grossman, L.; Moldave, K., editors. *Methods in Enzymology*. Vol. 101. 1983. p. 98-122.
- Beckmann JS, Weber JL. Survey of human and rat microsatellites. *Genomics* 1992;12:627–631. [PubMed: 1572635]

- Beerli P, Hotz H, Uzzell T. Geologically dated sea barriers calibrate a protein clock for Aegean water frogs. *Evolution* 1996;50:1676–1687.
- Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* 2006;34:1512–1521. [PubMed: 16554555]
- Buratti E, Baralle FE. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 2004;24:10505–10514. [PubMed: 15572659]
- Burch JBE, Davis DL, Haas NB. Chicken repeat 1 elements contain a *pol*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc Natl Acad Sci USA* 1993;90:8199–8203. [PubMed: 8396264]
- Burwinkel B, Kilmann MW. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J Mol Biol* 1998;277:513–517. [PubMed: 9533876]
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 1979;18:5294–5299. [PubMed: 518835]
- Deininger PL, Batzer MA. Mammalian retroelements. *Genome Res* 2002;12:1455–1465. [PubMed: 12368238]
- Deininger PL, Moran JV, Batzer A, Kazazian HH Jr. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 2003;13:651–658. [PubMed: 14638329]
- Edelmann W, Kröger B, Goller M, Horak I. A recombination hotspot in the LTR of a mouse retrotransposon identified in a *vitro* system. *Cell* 1989;57:937–946. [PubMed: 2544295]
- Eickbush, TH. Origin and evolutionary relationships of retroelements. In: Morse, SS., editor. *The evolutionary biology of viruses*. Raven Press; New York: 1994. p. 121–157.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
- Feschotte C. Transposable elements and the evolution of regulatory networks. *Nature Rev Genet* 2008;9:397–405. [PubMed: 18368054]
- Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell* 2002;110:315–325. [PubMed: 12176319]
- Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol Cell Biol* 2005;25:7780–7795. [PubMed: 16107723]
- Haas NB, Grabowsky JM, North J, Moran JV, Kazazian HH, Burch JB. Subfamilies of CR1 non-LTR retrotransposons have different 5' UTR sequences but are otherwise conserved. *Gene* 2001;265:175–183. [PubMed: 11255020]
- Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–174. [PubMed: 3934395]
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
- John JS, Cotter JP, Quinn TW. A recent chicken repeat 1 retrotransposition confirms the Coscoroba-Cape Barren goose clade. *Mol Phyl Evol* 2005;37:83–90.
- Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 1992;8:275–282. [PubMed: 1633570]
- Kaiser VB, van Tuinen M, Ellgren H. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol* 2007;24:338–347. [PubMed: 17077154]
- Kajikawa M, Ohshima K, Okada N. Determination of the entire sequence of turtle CR1: The first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol Biol Evol* 1997;14:1206–1217. [PubMed: 9402732]
- Kapitanov VV, Jurka J. The Esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol Biol Evol* 2003;20:38–46. [PubMed: 12519904]
- Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science* 2004;303:1626–1632. [PubMed: 15016989]

- Kazazian HH Jr, Goodier JL. LINE drive: retrotransposition and genome instability. *Cell* 2002;110:277–280. [PubMed: 12176313]
- Kordiš D, Lovšin N, Gubenšek F. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst Biol* 2006;55:886–901. [PubMed: 17345671]
- Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 200910.1093/bioinformatics/btp187
- Luan DD, Eickbush TH. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol* 1995;15:3882–3891. [PubMed: 7540721]
- Lymberakis P, Poulakakis N, Manthou G, Tsigenopoulos CS, Magoulas A, Mylonas M. Mitochondrial phylogeography of *Rana (Pelophylax)* populations in the eastern Mediterranean region. *Mol Phy Evol* 2007;44:115–125.
- Matsuba C, Merilä J. Isolation and characterization of 145 polymorphic microsatellite loci for the common frog (*Rana temporaria*). *Mol Ecol Resour* 2009;9:555–562.
- McDonald JHM, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 1991;351:652–654. [PubMed: 1904993]
- Minghetti PP, Ruffner DE, Kuang WJ, Dennison OE, Hawkins JW, Beattie WG, Dugaiczky A. Molecular structure of the human albumin gene is revealed by nucleotide sequence within q11-22 of chromosome 4. *J Biol Chem* 1986;261:6747–6757. [PubMed: 3009475]
- Nadir E, Margalit H, Gallily T, Ben-Sasson SA. Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 1996;93:6470–6475. [PubMed: 8692839]
- Novikova, O.; Šlivniška, E.; Fet, V.; Settele, J.; Blinov, A.; Woyciechowski, M. CR1 clade of non-LTR retrotransposons from *Maculinea* butterflies (Lepidoptera: Lycaenidae): evidence from recent horizontal transmission; *BMC Evolutionary Biology*. 2007. p. 93 (<http://www.biomedcentral.com/1471-2148/7/93>)
- Ostertag, EM.; Kazazian, HH, Jr. *Encyclopedia of Life Sciences*. John Wiley & Sons; 2005. Retrotransposition and human disorders; p. 1-6.
- Piriyaongsa J, Jordan IK. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PloS ONE* 2007;2:e203. [PubMed: 17301878]
- Piriyaongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 2008;14:814–821. [PubMed: 18367716]
- Piriyaongsa J, Mariño-Ramírez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics* 2007;176:1323–1337. [PubMed: 17435244]
- Plötner J. Genetic diversity in mitochondrial 12S rDNA of western Palearctic water frogs (Anura, Ranidae) and implications for their systematics. *J Zool Syst Evol Res* 1998;36:191–201.
- Plötner, J. *Von Märtyrern der Wissenschaft zur biologischen Sensation*. Laurenti Verlag; Bielefeld: 2005. Die westpaläarktischen Wasserfrösche.
- Plötner J, Ohst T. New hypotheses on the systematics of the western Palearctic water frog complex (Anura, Ranidae). *Mitt Mus Nat kd Berl, Zool Reihe* 2001;77:5–21.
- Plötner J, Ohst T, Böhme G, Schreiber R. Divergence in the mitochondrial DNA of Near Eastern water frogs with special reference to the systematic status of Cypriote and Anatolian populations (Anura, Ranidae). *Amphibia-Reptilia* 2001;22:397–412.
- Plötner, J.; Köhler, F.; Uzzell, T.; Beerli, P. Molecular systematics of amphibians. In: Heatwole, H.; Tyler, M., editors. *Amphibian Biology*. Vol. 7. *Phylogeny and Systematics*.: Surrey Beatty & Sons, Chipping Norton; 2007. p. 2672-2756.
- Plötner, J.; Uzzell, T.; Beerli, P.; Akin, Ç.; Bilgin, CC.; Haefeli, C.; Ohst, T.; Köhler, F.; Schreiber, R.; Guex, G-D.; Litvinchuk, AN.; Westaway, R.; Reyher, H-U.; Hotz, H. Genetic divergence and evolution of reproductive isolation in eastern Mediterranean water frogs. In: Glaubrecht, M.; Schneider, H., editors. *Evolution in action. Case studies in adaptive radiation and the origin of biodiversity*. Special volume from the SPP 1127 “Radiations – Genesis of Biological diversity” of the DFG. Springer; Heidelberg, Berlin: in press
- Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics* 1998;14:817–818. [PubMed: 9918953]

- Poulter R, Butler M, Ormandy J. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. *Gene* 1999;227:169–179. [PubMed: 10023050]
- Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Rev Genet* 2006;7:211–221. [PubMed: 16485020]
- Rozen, S.; Skaletsky, HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S.; Misener, S., editors. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press; Totowa, NJ: 2000. p. 365–386.
- Santucci F, Nascetti G, Bullini L. Hybrid zones between two genetically differentiated forms of the pond frog *Rana lessonae* in southern Italy. *J Evol Biol* 1996;9:429–450.
- Segal Y, Peissel B, Renieri A, de Marchi M, Ballabio A, Pei Y, Zhou J. LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. *Am J Hum Genet* 1999;64:62–69. [PubMed: 9915944]
- Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human genomic deletions mediated by recombination between *Alu* elements. *Am J Hum Genet* 2006;79:41–53. [PubMed: 16773564]
- Shedlock AM. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst Biology* 2006;55:902–911.
- Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci USA* 2007;104:2767–2772. [PubMed: 17307883]
- Silva R, Burch JBE. Evidence that chicken CR1 elements represent a novel family of retrotransposons. *Mol Cell Biology* 1989;9:3563–3566.
- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nature Rev Genet* 2007;8:272–285. [PubMed: 17363976]
- Stumph WE, Kristo P, Tsai MJ, O'Malley BW. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Res* 1981;9:5383–5398. [PubMed: 7301590]
- Stumph WE, Baez M, Beattie WG, Tsai MJ, O'Malley BW. Characterization of deoxyribonucleic acid sequences at the 5' and 3' borders of the 100 kilobase pair ovalbumin gene domain. *Biochemistry* 1983;22:306–315. [PubMed: 6297544]
- Stumph WE, Hodgson CP, Tsai MJ, O'Malley BW. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. *Proc Natl Acad Sci USA* 1984;81:6667–6671. [PubMed: 6593723]
- Swofford, DL. *Phylogenetic Analysis Using Parsimony (*and other Methods)*. Sinauer Associates; Sunderland, Massachusetts: 2003. PAUP*. Version 4
- Symer DE, Conelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* 2002;110:327–338. [PubMed: 12176320]
- Tabaska JE, Zhang MQ. Detection of polyadenylation signals in human DNA sequences. *Gene* 1999;231:77–86. [PubMed: 10231571]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. [PubMed: 17488738]
- Treplin S, Tiedemann R. Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picanthartes*) to crows and ravens (Corvidae). *Mol Phyl Evol* 2007;43:328–337.
- Uzzell T. Immunological relationship of western Palearctic water frogs (Salientia: Ranidae). *Amphibia-Reptilia* 1982;3:135–143.
- Watanabe M, Nikaido M, Tsuda TT, Inoko H, Mindell DP, Murata K, Okada N. The rise and fall of the CR1 subfamily in the lineage leading to penguins. *Gene* 2006;365:57–66. [PubMed: 16368202]
- Zaratiegui M, Irvine DV, Martienssen RA. Noncoding RNAs and gene silencing. *Cell* 2007;128:763–776. [PubMed: 17320512]

Figure 1.

A: Diagrammatic representation of the serum albumin intron-1 of western Palearctic water frogs and the inserted CR1 retrotransposon. B: Primary structure of the 5' end (pos. 287–317) and the 3' UTR (pos. 734–815) of *Rana*CR1 obtained from a Central European *R. ridibunda* (population Lebus, 16-DE). TSD: target site duplication.

Figure 2. Phylogenetic relationships (Maximum-Likelihood tree) of western Palearctic water frogs based on the SAI-1 and *RanaCR1* sequences. Bootstrap values greater than 50% (1000 replicates) are given above branches. The units of the scale bar are the expected mutations per site.

Table 1

Identification, origin (locality, coordinates), and number (N) of western Palearctic water frogs studied, together with EMBL accession numbers for the nucleotide sequences. Country abbreviations are given in squared brackets.

Species	Origin		Locality	Latitude	Longitude	N	Accession No.
	Country						
<i>R. bedriagae</i>	Syria [SY]	01	As Suwayda	32°43' N	36°33' E		
<i>R. cf. bedriagae</i>	Greece [GR]	02	Chios	38°28' N	25°56' E	1	FN432372
	Turkey [TR]	03	Akcapinar	37°02' N	28°22' E	1	FN432373
<i>R. cretensis</i>	Greece [GR]	04	Demati (Crete)	35°02' N	25°17' E	1	FN432374, FN432375
	Greece [GR]	05	Skinias (Crete)	35°04' N	25°19' E	1	FN234376
<i>R. epeirotica</i>	Greece [GR]	06	Igoumenitsa	39°36' N	20°03' E	1	FN432369
	Greece [GR]	07	Lechena	37°55' N	21°03' E	1	FN432370
<i>R. lessonae</i>	Italy [IT]	08	Carbonare	45°56' N	11°13' E	1	FN432383
	Germany [DE]	09	Muhlenbeck	52°40' N	13°21' E	1	FN432384
	Germany [DE]	10	Wernitz	52°34' N	12°55' E	1	FN432385
	Italy [IT]	11	Metaponto	40°22' N	16°49' E	1	FN432381
	Italy [IT]	12	Tarsia	39°37' N	16°16' E	1	FN432382
<i>R. perezi</i>	France [FR]	13	Ligagneau	43°32' N	04°45' E	1	FN432377
	Portugal [PT]	14	Monchique	37°19' N	08°33' W	1	FN432378
<i>R. ridibunda</i>	Kazakhstan [KZ]	15	Atyrau	47°06' N	51°55' E	4	FN432363, FN432371
	Germany [DE]	16	Lebus near Frankfurt/Oder	52°25' N	14°32' E	1	FN432364
	Poland [PL]	17	Poznan	52°25' N	16°53' E	1	FN432365
	Greece [GR]	18	Nea Manolada	37°42' N	21°24' E	1	FN432367
	Greece [GR]	19	Kavasilas	37°52' N	21°16' E	1	FN432366
<i>R. saharica</i>	Algeria [DZ]	20	El Golea	30°34' N	02°53' E	1	FN432379
	Algeria [DZ]	21	Abadla	31°01' N	02°44' E	1	FN432380

Table 2

Primers designed for amplification and cycle sequencing. Ex: Exon, Int: Intron, F: forward, R: reverse.

Name	Sequence (5'→3')	Location
Ex1-F2	ACTCTGATTGTCTGTTTATTC	exon 1, position -73
Ex1-F5	ATAACAACGAGTCCAGACCAC	exon 1, position -46
Ex2-R2	CTGCCTTTACAATATCGTTTAT	exon 2, position +33
Ex2-R3	AATTTTTCAACAGCTGGTTTCC	exon 2, position +58
Int-F1	CTCTTTTAATAACTGAGGAAGATGAGG	intron, position 85
Int-F6	TGAACTTATTATCTCTGGAGAAGAG	intron, position 385
Int-F7	GGGCACCTTAAAGACCACAA	intron, position 663
Int-R1.1	CCCCCTCAATCGTCTCTTCT	intron, position 422
Int-R1.0	CCGCTCTGACAGTAAAGAACCCC	intron, position 585
Int-R5	ATACCCCTGTATGTTGTGGTCTTTA	intron, position 694
Int-R1	GATTTTTATGGCTGTTGTTGTGTC	intron, position 1021

Table 3

Size and GC content of albumin intron-1 (SAI-1) and *Rana*CR1 sequences of water frogs specified for the carboxy-terminal region (CTR) and the 3'UTR, and the repeat number and sequence of the microsatellite integrated into the 3' UTR. Mean values and standard deviations (SD) were calculated for WPWF species on the basis of complete sequences only; unusual values caused by large deletions were not considered (in parentheses).

Species	ID	Allele	SAI-1 Size [bp]	GC [%]	<i>Rana</i> CR1 Size [bp]	GC [%]	CTR Size [bp]	GC [%]	3'UTR Size [bp]	GC [%]	Microsat
<i>R. bedriagae</i>	01-SY		535	34.4	491	40.3	386	40.9	80	37.5	(CA) ₆
<i>R. cf. bedriagae</i>	02-GR		560	33.1	489	39.9	386	40.1	78	38.5	(CA) ₅
	03-TR		543*	33.2	489	39.9	385	40.3	79	38.0	(CA) ₅
<i>R. cretensis</i>	04-GR	a	569	34.3	478	39.3	375	39.7	78	37.2	(CA) ₅
		b	569	34.1	478	39.1	375	39.5	78	37.2	(CA) ₅
	05-GR		565*	35.1	478	39.5	375	40.0	78	37.2	(CA) ₅
<i>R. epeirotica</i>	06-GR		562	33.8	487	41.0	385	41.5	77	39.0	(CA) ₆
	07-GR		561	33.9	487	41.0	385	41.5	77	39.0	(CA) ₆
<i>R. lessonae</i>	08-IT		426	34.0	92	41.4	(16)	(50.0)	76	39.5	CAGGCACA
	09-DE		389*	33.9	92	41.4	(16)	(50.0)	76	39.5	CAGGCACA
	10-DE		415*	33.7	92	41.4	(16)	(50.0)	76	39.5	CAGGCACA
	11-IT		424	33.3	84	42.8	(16)	(50.0)	68	41.1	CAGGCACA
	12-IT		426	33.6	84	42.8	(16)	(50.0)	68	41.1	CAGGCACA
<i>R. perezi</i>	13-FR		571	33.2	503	41.2	393	41.7	85	38.8	(CA) ₈
	14-PT		332*	32.6	499	41.0	393	41.7	81	38.3	(CA) ₆
<i>R. ridibunda</i> (type locality)	15-KZ	a	560	32.9	489	39.4	386	39.9	78	38.5	(CA) ₅
		b	560	33.4	491	40.8	386	41.7	80	37.5	(CA) ₆
<i>R. ridibunda</i> (CE)	16-DE		560	33.7	492	40.9	385	41.6	82	37.8	(CA) ₇
	17-PL		559	33.8	493	40.7	386	41.5	82	37.8	(CA) ₇
<i>R. ridibunda</i> (Balkan)	18-GR		497	33.6	430	41.9	386	41.8	(19)	(47.4)	(CA) ₆
	19-GR		496	33.7	430	41.9	386	41.8	(19)	(47.4)	(CA) ₆
<i>R. saharica</i>	20-DZ		573	33.1	499	40.0	393	40.7	81	38.3	(CA) ₆
	21-DZ		573	32.9	495	40.5	390	41.3	79	38.3	(CA) ₅
<i>R. nigromaculata</i>	22-KR		540	35.2	31	45.2	-	-	-	-	-
		Mean ±SD	532.3 ±54.17	33.6 ±0.56	397.0 ±167.33	40.8 ±1.03	385.3 ±5.54	41.0 ±0.82	78.0 ±4.02	38.5 ±1.13	-

ID: individual sequence identification number, combining the country abbreviation with the locality number (see Table 1). KR: North Korea. CE: Central Europe.

* Sequence incomplete.