Genetic divergence and evolutionary times: calibrating a protein clock for South-European *Stenasellus* species (Crustacea, Isopoda)

Valerio Ketmaier, Roberto Argano, Marina Cobolli and Elvira De Matthaeis*

SUMMARY

We studied genetic divergence in a group of exclusively stygobiont isopods of the family Stenasellidae. In particular, we assessed evolutionary relationships among several populations of *Stenasellus racovitzai* and *Stenasellus virei*. To place this study in a phylogenetic context, we used another species of *Stenasellus*, *S. assorgiai*, as an outgroup. *S. racovitzai* occurs in Corsica, Sardinia and in the fossil islands of the Tuscan Archipelago, while *S. virei* is a polytypic species widely distributed in the central France and Pyrenean area. This vicariant distribution is believed to be the result of the disjunction of the Sardinia-Corsica microplate from the Pyrenean region and its subsequent rotation. Since geological data provide time estimates for these events, we can use the genetic distance data to calibrate a molecular clock for this group of stygobiont isopods. The calibration of the molecular clock reveals a roughly linear relationship ($r = 0.753$) between the genetic distances and absolute divergence times, with a mean divergence rate (19.269 Myr/DNei) different from those previously reported in the literature and provides an opportunity to shed some light on the evolutionary scenarios of other *Stenasellus* species.

INTRODUCTION

Since Darwin and Wallace, biologists have emphasised the crucial importance of geographical isolation in speciation and in the subsequent adaptation and differentiation of species. The present range of several taxa may be seen as a combination of palaeogeographical events and dispersal phenomena. Some organisms have a poor dispersal capability and thus many barriers would constitute insurmountable obstacles to their dispersal. For example, stygobiont animals are strictly bound to the continental subterranean freshwaters, their potential for active or passive dispersal being limited to phreatic systems. Therefore, it can be hypothesised that their distribution reflects more the palaeogeographical and palaeogeological history of the occupied land masses than the active dispersal processes.

One of the most hotly debated topics in evolutionary studies is the possibility to estimate divergence times from genetic and/or morphological data.

* Dipartimento di Biologia Animale e dell’Uomo, Università di Roma «La Sapienza». Viale dell’Università 32, I - 00185 Roma, Italy. E-mail: ketmaier@axrma.uniroma1.it.
The major constraint to the use of genetic divergence data for the dating of cladogenetic events is the scarcity of related fossil records and their uncertain interpretation for calibration of a molecular clock. Nei (1975) proposed a linear relationship between divergence times and the degree of genetic differentiation. Several studies (for a review, see Sbordoni et al., 1990) have demonstrated that Nei's equation can be used for relatively small genetic distances (D<1), while for higher D values a calibration is needed. Sarich (1977) observed differences in divergence rates among proteins and recognised groups of fast and slow evolving genes. Other studies have demonstrated that a standardised clock for all genes and species does not exist (Scherer, 1990; Gillespie, 1991). Wilson et al. (1977) developed a relative rate test to investigate the existence of a universal clock. However, there is much evidence to reject the hypothesis of a universal clock across very different lineages; the notion of a local clock with lineage-specific rates seems to be more realistic (Nei, 1987). Testing the existence of a molecular clock is of crucial importance in evolutionary studies, but at least one independently timed event has to be used (Busack, 1986). A molecular clock with known relationships among the taxa under study and with well-dated isolation times may provide an accurate estimate, especially when cladogenetic events can be associated with specific well-dated biogeographical scenarios (Caccone et al., 1994; 1997).

According to Alvarez (1972) and Bellon et al. (1977), Corsica and Sardinia separated as a single land mass from the Iberian Peninsula in the Miocene (about 29-27 Mya). After the disjunction, a rotation brought the two islands to their present positions. The separation of Corsica from Northern Sardinia may have begun about 15 Mya and was certainly completed by 9 Mya. The interaction between the Corsica-Sardinia microplate and the Apennines (then being formed) caused the emergence of the Tuscan Archipelago, including the fossil islands presently incorporated in the mainland (e.g. Monte Massoncello, Monti dell’Uccellina, etc.) (from 9 to 7 Mya) (Lanza, 1984). Several taxa are present on each of these land masses (and only on these) with closely related species. Since the potential for dispersal of many of the species is very low or completely absent (especially for cave organisms), it is reasonable to link the cladogenetic events in these lineages with the palaeogeographical history of the area.

Several genetic studies on phylogenetically distinct species showing this vicariant distribution, such as newts and troglobitic beetles, highlighted these tectonic events as an unique model to test the hypothesis of a clock-like behaviour in the increase of the genetic divergence (Caccone et al., 1994; 1997; Sbordoni et al., 1982; 1985; 1990; 1992).
In previous papers (Messana et al., 1995; Ketmaier et al., 1999b), we assessed the level of genetic differentiation among several populations of the three Italian species of the genus *Stenasellus* (*S. racovitzai, S. nuragicus, S. assorgiai*), which belongs to the exclusively stygobiont family Stenasellidae. Among the South-European species of *Stenasellus, S. racovitzai* shows a close morphological affinity with the Pyrenean species *S. virei* (Magniez, 1974). *S. racovitzai* is presently known from two localities in Tuscany, from the water-bearing strata of the hydrographic systems of the Golo River in Corsica and the Rio di Quirra basin in Southeast Sardinia and from an artificial well close to Porto Torres in Northwest Sardinia (Argano et al., 1998). *S. virei* is a polytypic species widely distributed in the Pyrenean area and central France. This vicariant distribution is a remarkable example of allopatric speciation which can be explained in the context of the previously discussed tectonic events.

The aims of this study were:
1. to determine the levels of genetic divergence within and between *S. racovitzai* and *S. virei*;
2. to test the hypothesis of a clock-like nature of the increase of genetic divergence in this group;
3. to calibrate a molecular clock for the genus *Stenasellus* using the palaeogeographical absolute divergence times.

**MATERIALS AND METHODS**

**Collecting sites and study samples**

Eight populations of *Stenasellus* (identified with a three-letter code) were tested for genetic divergence and polymorphisms. The populations of *S. racovitzai* were a part of those analysed in Ketmaier et al. (1999a):
- DAN: from the type locality (Danese cave, Grosseto, Tuscany);
- UCC: from a small cave (Fontanile dei Cavalleggeri) on the Monti dell’Uccellina slope near the sea (Tuscany);
- QUI: from the water-bearing strata of the Rio di Quirra, Sardinia;
- COR: from the water-bearing strata of the Golo River, Corsica.

Two populations of *S. virei virei* and two populations of *S. virei hussoni* were studied:
- VIR: *S. virei virei*, from the water-bearing strata of the Dordou River, S. Izaires, Albi (France);
- TAR: *S. virei virei*, from the water-bearing strata of the Tarn River, Les-cure, Albi (France);
- HUS: *S. virei hussoni*, from the Estelas cave, Pyrenees (France);
- PEY: *S. virei hussoni*, from the Peyrot cave, Pyrenees (France).

Moreover, for the phylogenetic analyses we used a population of *S. assorgiai* (PIT) from the type locality (Pitzu 'e Crobisi cave, Cagliari, Sardinia) as an outgroup.

The collecting sites are shown in Fig.1.

Live specimens were transported to the laboratory and frozen at -80°C.

---

**Electrophoretic procedures and data analysis**

Horizontal electrophoresis on 12% starch gel was performed with crude homogenates in TRIS HCl 0.05 M pH 7.5 from each whole specimen. Eleven enzymes, encoded by 15 structural gene loci, were examined: Acid phosphatase (*Acph*, E.C.No.: 3.1.3.2); Alkaline phosphatase (*Aph*, E.C.No.: 3.1.3.1); Esterase (*Est-1, Est-2, Est-3*, E.C.No.: 3.1.1.1); Isocitrate dehydrogenase (*Idh*, E.C.No.: 1.1.1.42); Lactate dehydrogenase (*Ldh*, E.C.No.: 1.1.1.27); Malate dehydrogenase (*Mdh*, E.C.No.: 1.1.1.37); Nothing dehydrogenase (*No-dh*, E.C.No.: 1.6.99.1); Peptidase (*Pep-1, Pep-2*, E.C.No.: 3.4.11); Phosphoglucomutase (*Pgm*, E.C.No.: 2.7.5.1); Phos-
phoexose isomerase (Phi, E.C.No.: 5.3.1.9); Tetrazolium oxidase (To, E.C.No.: 1.15.1.1). The procedures used are those in Messana et al. (1995).

The genetic relationships between populations and species were estimated by genetic distance values (D) calculated according to Nei (1978) on the basis of the allele frequencies at 15 common loci and were represented by a dendrogram plotted with the UPGMA clustering method (Sneath & Sokal, 1973). To test phylogenetic hypotheses, we analysed our data set by two different methods: Maximum Likelihood (ML; programme CONTML in Phylip 3.5, Felsenstein, 1993) and Neighbor-Joining (NJ; programme NEIGHBOR in Phylip 3.5, Felsenstein, 1993). The robustness of the trees was tested by the bootstrap method (programme SEQBOOT in Phylip 3.5, Felsenstein, 1993) with 1000 replications for both phylogenetic methods.

The genetic variability of the populations was estimated by He (expected mean heterozygosity under Hardy-Weinberg equilibrium), Ho (observed heterozygosity), P (proportion of polymorphic loci according to the criterion of the second most common allele being at least 1%) and A (mean number of alleles per locus).

To assess evolutionary rates, we performed a regression of the genetic distances against the absolute isolation times deduced by palaeomagnetic, stratigraphic and geomorphological data (Alvarez, 1972; 1974; Alvarez et al., 1973; Ambrosetti et al., 1979; Bellon et al., 1977; Bonin et al., 1979; Orsini et al., 1980; Cherchi & Montadert, 1982; Esu & Kotsakis, 1983).

Statistical analyses were performed with the BIOSYS-1 programme (Swofford & Selander, 1981), STATISTICA 4.5 for Windows and the PHYLIP 3.5 package (Felsenstein, 1993).

RESULTS

Fifteen inferred loci were consistently scored; three (No-dh, Pep-2 and To) were monomorphic in all study populations, while the remaining twelve were polymorphic in at least one population. Seven fixed alternative alleles were scored among all study populations. Allele frequencies are available from the first author upon request.

A wide range of D values was found, from 0.009 (VIR vs TAR) to 3.030 (QUI vs PIT). For S. racovitzai, the genetic distance between UCC and DAN was 0.175, but the value almost reached 1 when the Tuscan (UCC, DAN) populations were compared with the Sardinian (QUI) and Corsican (COR) ones and when the two island populations were compared (QUI vs COR = 0.952). The genetic distance data are reported in Table 1, while the genetic relationships among populations are represented in the UPGMA dendrogram of Fig. 2. ML and NJ gave trees with identical to-
polopies which are summarised in Fig. 3; both trees were rather robust in
terms of bootstrap values.

The genetic variability of the populations (Table 2), expressed as Ho,
varied from 0.000 (VIR; PIT) to 0.139 (HUS).

The graphical result of the regression analysis of the genetic distances
against the absolute times is shown in Fig. 4. The regression line was Myr =
0.234 + 19.262 DNei and the correlation coefficient was r = 0.753.

Table 1 - Genetic distance values D (Nei, 1978) between study Stenellas populations.

<table>
<thead>
<tr>
<th>Pop.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - UCC</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 - DAN</td>
<td>0.175</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 - QUI</td>
<td>1.138</td>
<td>1.128</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 - COR</td>
<td>0.954</td>
<td>0.686</td>
<td>0.952</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 - HUS</td>
<td>1.075</td>
<td>0.971</td>
<td>0.947</td>
<td>1.136</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 - PEY</td>
<td>0.994</td>
<td>0.829</td>
<td>0.817</td>
<td>1.025</td>
<td>0.082</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 - VIR</td>
<td>1.081</td>
<td>1.062</td>
<td>1.445</td>
<td>1.383</td>
<td>0.291</td>
<td>0.281</td>
<td>****</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 - TAR</td>
<td>1.165</td>
<td>1.132</td>
<td>1.438</td>
<td>1.474</td>
<td>0.248</td>
<td>0.271</td>
<td>0.009</td>
<td>****</td>
<td></td>
</tr>
<tr>
<td>9 - PIT</td>
<td>1.689</td>
<td>1.607</td>
<td>3.050</td>
<td>1.584</td>
<td>1.752</td>
<td>1.576</td>
<td>1.322</td>
<td>1.373</td>
<td>****</td>
</tr>
</tbody>
</table>

Table 2 - Variability estimates for the study Stenellas populations (Standard Errors in parentheses).

<table>
<thead>
<tr>
<th>Population</th>
<th>Mean sample size per locus</th>
<th>Mean number of alleles per locus (A)</th>
<th>Proportion of polymorphic loci* (P)</th>
<th>Mean heterozigosity, Direct count (Hd)</th>
<th>Mean heterozigosity HDYWBG expected** (H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - UCC</td>
<td>22.7 (0.3)</td>
<td>1.3 (0.1)</td>
<td>33.3</td>
<td>0.042 (0.020)</td>
<td>0.047 (0.024)</td>
</tr>
<tr>
<td>2 - DAN</td>
<td>16.0 (0.0)</td>
<td>1.3 (0.2)</td>
<td>20.0</td>
<td>0.067 (0.054)</td>
<td>0.047 (0.035)</td>
</tr>
<tr>
<td>3 - QUI</td>
<td>24.7 (0.3)</td>
<td>1.4 (0.1)</td>
<td>40.0</td>
<td>0.092 (0.054)</td>
<td>0.119 (0.049)</td>
</tr>
<tr>
<td>4 - COR</td>
<td>5.9 (0.1)</td>
<td>1.2 (0.1)</td>
<td>20.0</td>
<td>0.033 (0.018)</td>
<td>0.049 (0.030)</td>
</tr>
<tr>
<td>5 - HUS</td>
<td>14.3 (0.7)</td>
<td>1.7 (0.2)</td>
<td>53.3</td>
<td>0.139 (0.038)</td>
<td>0.196 (0.054)</td>
</tr>
<tr>
<td>6 - PEY</td>
<td>19.0 (0.5)</td>
<td>1.4 (0.1)</td>
<td>40.0</td>
<td>0.079 (0.031)</td>
<td>0.116 (0.042)</td>
</tr>
<tr>
<td>7 - VIR</td>
<td>14.3 (0.7)</td>
<td>1.0 (0.0)</td>
<td>0.00</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>8 - TAR</td>
<td>13.4 (0.3)</td>
<td>1.4 (0.1)</td>
<td>40.0</td>
<td>0.033 (0.017)</td>
<td>0.087 (0.038)</td>
</tr>
<tr>
<td>9 - PIT</td>
<td>8.7 (0.2)</td>
<td>1.0 (0.0)</td>
<td>0.00</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
</tbody>
</table>

* A locus is considered polymorphic if the frequency of the most common allele does not exceed 0.99.
** Unbiased estimate (see Nei, 1978)

DISCUSSION

We found a certain degree of divergence between the two subspecies
of S. virei, which could indicate that a process of differentiation is taking
place within the species. However, further study of a larger number of
populations is needed to elucidate the systematics of S. virei.

Preliminary genetic data for S. racovitzai (Messana et al., 1995) clearly
showed the existence of two well-differentiated species from Tuscany and
Sardinia. In the present study, we also analysed the Corsican population of *S. racovitzai*. Its mean degree of genetic differentiation from the Tuscan and Sardinian populations strongly supports the hypothesis of three well-differentiated species on the three land masses. The lack of morphological differences among cave species that are well differentiated genetically is a common phenomenon (Sbordoni, 1982; Cobolli Sbordoni et al., 1990). This seems to be related to the uniformity of selective pressures acting in cave environments, which leads to high levels of morphological similarity due to parallelism and/or convergence. In a companion paper (Ketmaier et al., 1999 b) we have emphasized that what we presently call *S. racovitzai* is, in fact, a species complex. This strongly supports the hypothesis that the cladogenetic events within this lineage are related to the palaeogeographical connection between Sardinia and Corsica, as well as between Corsica and the Tuscan Archipelago (including the fossil islands).

The ML and NJ analyses indicate an evolutionary scenario based on a vicariance model. *S. racovitzai* and *S. virei* are sister species, and within the *S. racovitzai* complex, there is a close relationship between the Corsican and Tuscan species. This finding is consistent with the palaeogeography of the study area, since Corsica and Tuscany were in contact even after the connection between Corsica and Sardinia disappeared. Thus we believe that a plausible evolutionary scenario leading to the present distribution of the *Stenasellus* species under study can be based on a vicariance model which assumes that an old, widely distributed species became subdivided during the above-mentioned palaeogeographical events.

Fig. 2 – UPGMA dendrogram constructed on the basis of the observed genetic distances D (Nei, 1978).
Fig. 3 – Majority rule consensus tree obtained by ML and NJ bootstrap analyses. Circled nodes include bootstrap percentages of 1000 replications for ML and NJ (first and second value, respectively). Bootstrap values are shown only for nodes for which the two phylogenetic methods had a bootstrap support of 70% or greater.

Fig. 4 – Genetic distances \( D \) (Nei, 1978) versus absolute geological isolation times (Myr). Regression line: \( \text{Myr} = 0.234 + 19.262 \cdot D_{\text{Nei}} \); correlation coefficient \( r = 0.753 \).

Interestingly enough, the mean level of genetic differentiation shown by our *Stenasellus* species of the same order of magnitude of that already found among species of the newts *Euproctus*, distributed in Corsica, Sardinia and Pyrenees (Sbordoni et al., 1990).

The peculiar ecological features of Stenasellidae and the ancient origin of the family itself (Magniez, 1974; 1981) allow us to overcome the kinds
of problems, e.g. severe bottlenecks, that could bias the calibration of the molecular clock. The idea of a clock-like nature of genetic divergence assumes a constant rate of genomic change through time. However, classic models of population genetics predict a temporary loss of variability and an increase of genetic divergence by random genetic drift in populations experiencing a severe bottleneck. Thus in the presence of bottleneck effects, the divergence rate will be overestimated (Nei, 1987). Organisms bound to phreatic or cave water systems have very low (if any) probability of accidental introduction to, or active migration through, unsuitable ecological areas. Heterozygosity data for our island populations do not indicate any loss of genetic variability by random genetic drift with respect to the continental populations. Hence these kinds of events can be excluded for our Stenasellus populations.

Although it has generally been recognised that there is a relationship between genetic divergence and time, the nature of this relationship is not very clear (Foley, 1987; Scherer, 1990). We found a strong correlation ($r = 0.753$) between the genetic distances and the absolute palaeogeographical time estimates, which supports the hypothesis of a linear relationship between geological time and genetic divergence.

The divergence rate calculated in this study is $19.269 \text{ Myr}/D_{\text{Nei}}$. Thus the rate of $14 \text{ Myr}/D_{\text{Nei}}$ estimated by Sarich (1977) seems to be too slow for our data and would produce misleading evolutionary scenarios. Caccone et al. (1994) stressed that «rate heterogeneities clearly exist at several genomic levels», electrophoretic data on Pyrenean and Sardinian species belonging to the beetle genus *Speonomus* are a further confirmation (Sbordoni et al., 1992), since genetic distance between them is almost twice the genetic distance between the same geographical comparison in *Stenasellus*. Indeed, various studies (Beerli et al., 1996; Hillis & Moritz, 1990) have demonstrated that the clock-like nature of genetic divergence is viable at the level of the same lineage but not across very different lineages. Sarich (1977) divided the set of commonly detected proteins into two main categories – fast and slow evolving loci – and pointed out that the former accumulate electrophoretically detectable substitutions at a rate ten times greater than the latter. In contrast, Thorpe (1989) suggested that there is no clear-cut bimodal distribution of the rates of protein evolution; instead, the differential rates occurring in a data set are probably the result of random sampling from a continuum rather than from well-separated classes. Beerli et al. (1996) emphasised, however, that these differences in the rate of genetic evolution do not falsify the concept of a molecular clock.

We divided our set of scored loci into slow and fast evolving loci according to the procedure proposed by Sarich (1977): dehydrogenase en-
zymes were considered as slow evolving loci, while the other enzymes were treated as fast evolving loci. This empirical assignment to the two groups was evaluated by recalibrating the molecular clock without the fast evolving loci. The mean divergence rate drops from 19.269 Myr/D_{Nei} to 12.561 Myr/D_{Nei}, confirming the existence of rate heterogeneities even at the level of enzymatic proteins. Several studies (Hillis & Moritz, 1990; Scherer, 1990; Gillespie, 1991; Wilson et al., 1987) have shown that multiple sub-sti-tutions in rapidly evolving parts of the genome (including regions for enzymatic proteins) can cause significant problems for large divergence times. However, these authors referred to a time scale longer than 100 Myr. Since our averaged protein clock spans a maximum of 30 Myr, it is reasonable to assume that it is not affected by these phenomena.

A particularly intriguing aspect of our study is the possibility to develop a mean divergence rate for the genus *Stenasellus* which can then be applied to populations and species for which no clear palaeogeographical scenarios are available. The above mentioned problems of rate heterogenei-ties among different lineages compel one to be cautious in extrapolating a divergence rate from one group to another. However, as shown by Beerli et al. (1996), extrapolations within the same genus are permitted.

Despite the degree of error implicit in the method, our calibration of a molecular clock for the genus *Stenasellus* will be used in the near future to place in a temporal frame the complex evolutionary scenarios (Ketmaier et al., 1999 a) presented by some endemic Sardinian species of this genus (i.e. *S. nuragicus, S. assorgiai*).

ACKNOWLEDGEMENTS

We wish to thank Dr. Claude Bou for assistance in collecting the French specimens and Dr. Giuseppe Messana for his continuous critical advices during this study.

This work was supported by MURST 40% and by Consiglio Nazionale delle Ricerche (CNR, Comitato Ambiente) grants. The data here presented are part of VK’s Ph.D. thesis project.

REFERENCES


