INTRODUCTION

A ‘centre of origin’ is the maximum centre of speciation from which species disperse to marginal locations. The ‘centre of origin’ concept has long dominated the interest of biogeographers (see Humphries & Parenti 1999, Morrone & Crisci 1995, Posadas et al. 2006) and a plethora of contrasting criteria to locate the centres of origin have been proposed (see Cain 1944 and Brown & Gibson 1983 for reviews). With the rise of vicariance biogeography, the ‘centre of origin’ quest has been dismissed as a false problem (Humphries & Parenti 1999): although dispersal cannot be refuted as a real phenomenon capable of producing disjunct distributions (e.g. Wilkinson 2003, Cook & Crisp 2005, Greve et al. 2005), it is considered as noise that obscures vicariance events, on which historical biogeography must be primarily founded (cf. Humphries & Parenti 1999, Fattorini 2008 for reviews). However, some attempts have been made to examine the ‘centre of origin’ concept by cladistic means with the identification of ‘ancestral areas’ (Bremer 1992, Ronquist 1994, 1995, Hausdorf 1998), and the problem of locating possible centres of origin has received renewed attention because of its possible implications in conservation biology (e.g. Fjeldsa 1994, Linder 1995, Solis & Gitzendanner 1999).

In this paper, we present a new approach to the centre of origin theory, which is based on a purely statistical analysis of matrix nestedness and which does not require phylogenetic information. In a perfectly nested distribution, species occurring in the area of interest are always present in a more species-rich area, whereas those not present never occur in a more depauperated one (Patterson & Atmar 1986). Nestedness is considered significant if patterns of species occurrence are closer to perfectly nested (ordered) distribution than expected by chance (Ulrich et al. 2009). However, no research has addressed the question of whether the order in a given matrix is widespread or concentrated in one or more particular sector(s). Finding that the order is concentrated in particular sectors of a matrix indicates that widespread species tend to be present in the species-richest area. A ‘centre of origin’ hypothesis, which assumes that species have a common place of origin corresponding to the richest area within the most ordered sector of the matrix and then spread through different localities (i.e. the rest of the matrix), appears as the most likely explanation to this pattern. Global distribution of siganid fishes is used to present this new approach, adding further momentum to the debate about the Indo-Malaysian-Philippine Archipelago centre of origin.

A new statistical technique able to localize nestedness within a matrix is used to reapproach the ‘centre of origin’ concept. Finding that nestedness is concentrated in a particular sector of a matrix indicates that widespread species tend to be present in the species-richest area. A possible explanation for such a situation is that the most ordered sector of the matrix is a centre of origin from which the species spread more or less independently to different areas. Areas far from the centre of origin display low species richness because only species with good dispersal abilities have colonized them.

In the past, it has been proposed to equate the centre of maximum diversity of a taxon with its supposed centre of origin, even if it does not include any of the species occurring in other areas (cf. Cain 1944). By contrast, in our nestedness approach, species identity is considered, and a centre of origin is recognized as the richest area that hosts the largest assemblage of species also occurring in other regions. Moreover, in our approach centres of origin are not identified on the basis of their high faunal nestedness, but on the basis of the fact that the species assemblages of other areas are subsets of the fauna of the centre of origin. Thus, we do not postulate a mechanism that increases or decreases nestedness within the centre of origin, but we propose a method to assess if faunas of peripheral areas are subsets of the fauna of a central area.

To illustrate our procedure, we used the global distribution of siganid fishes (Actinistia: Siganidae). The Siganidae is a monogeneric family of 28 species (Froese & Pauly 2009) with a number of characteristics that make them of particular biogeographical interest.

First, siganid taxonomy is relatively stable and species distributions are well documented.

Second, siganids have very similar ecological niches. They are primarily herbivorous, living on benthic algae. Although larvae are pelagic, larval life-span is only 3-4
weeks. This prevents the possibility of very long-dispersal across the sea in a short time. On the other hand, these organisms are sufficiently good dispersers to make unlikely that their current distribution is determined only by vicariance events (see Woodland 1983, 1990, 1999). The Mediterranean invasions of Siganus rivulatus and S. luridus (Ben-Tuvia 1964) is evidence of the importance of dispersal in determining siganid distribution.

Third, although phylogenetic information (Borsa et al. 2007, Kuriwa et al. 2007, Lemer et al. 2007) is insufficient for a detailed reconstruction of the biogeographical history of the family, it is nonetheless adequate to be compared with a biogeographical reconstruction based on purely distributional data.

Fourth, the choice of a monogeneric family with fewer than 30 acknowledged species allowed us to face distributional patterns likely to have evolved over evolutionary time with a worldwide number of species sufficiently large to develop the analysis, but not so large (as in other fish or invertebrate families) to render an analysis of individual species’ ranges excessively cumbersome.

Finally, high diversity in Indo-Malaysian-Philippine Archipelago (IMPA) has been recently evoked to support the centre of origin theory (Mora et al. 2003, Carpenter & Springer 2005). The distribution of the Siganidae throughout the whole tropical Indo-Pacific region, and their peak of species-richness in the Indo-West Pacific Ocean (Woodland 1990) represent a classical example of controversy about the centre of origin (see Briggs 2000 for a detailed review of competing hypotheses). Thus, we think that the use of siganids to illustrate our approach may add further momentum to the debate about the IMPA centre of origin.

MATERIALS AND METHODS

Data coding: species presence/absence was coded using marine regions delimited by major faunal boundaries affecting reef fish taxa, as reported by Bellwood & Wainwright (2002) (Fig. 1, Table I). Several authors pointed out the need to use ‘areas of endemism’, in biogeographical studies, in order to avoid the extreme incongruent patterns that result from mixing different ancestral biota (e.g. Harold & Mooi 1994, Morrone & Crisci 1995, Glasby & Alvarez 1999). Therefore, as in other biogeographical studies about marine fauna (van Soest & Hajdu 1997, Glasby & Alvarez 1999, Santini & Winterbottom 2002, Garrafoni et al. 2006) we adopted a system of regions which are likely to reflect natural boundaries.

Point locality records for each siganid species were primarily obtained from Woodland (1990), and updated using subsequent literature data (Rao & Devi 1996, Azzurro & Andaloro 2004, Dulcic & Pallaoro 2004, Randall & Kulbicki 2005, Kuriwa et al. 2007), and the online database Fishbase (Froese & Pauly 2009).

Statistical procedure: to assess whether the order (or disorder) in a certain binary matrix is distributed uniformly, or results from the combination of ordered and disordered regions within the matrix, we used the following approach:

a – We partitioned the complete matrix (here termed “C-Matrix”) into a certain number of submatrices corresponding to meaningful major biogeographical groups of areas.

b – We assessed nestedness of each submatrix using Z-values. Z-value is defined as:

\[ Z = \frac{(Ta - Tr)}{Sd} \]

where Ta is the actual temperature of the submatrix, Tr is the average temperature of a set of matrices obtained from randomization of that submatrix and Sd is the standard deviation of temperatures of the random matrices. Negative Z-values point to nestedness.

c – From the original C-matrix, we extracted a certain number of submatrices at random but with the same number of columns as for the submatrices corresponding to the biogeographical groups of areas. In particular, we extracted by random selection (without repetition) of columns (i.e. areas) as many sets of 1000 matrices as were the submatrices identified using biogeographical criteria. For example, if we have a complete matrix of 25 columns, from which three submatrices are identified on a biogeographical basis, comprising 7, 6 and 5 columns (areas) respectively (7 columns are left because they are not included in any biogeographical group of areas), we produce 1000 matrices...
of 7 columns, 1000 matrices of 6 columns, and 1000 matrices of 5 columns by extracting at random 7, 6 and 5 columns from the complete matrix (in our real data set we have a C-matrix of 24 columns which is partitioned into three matrices with the same number (8) of columns, so we have three sets of 1000 matrices of 8 columns).

d – For each of these matrices constructed by sampling columns from the C-matrix we calculated the respective Z-value using the same formula reported at point b.

e – Then, for each set of 1000 matrices, we calculated the percentage of random matrices showing a Z-value smaller than or equal to that of the corresponding ‘biogeographical’ submatrix and used this value as a p-value to test the ‘relative nestedness’ of submatrices. Considering the example reported above, we calculated the Z-value of each of the 1000 matrices of 7 columns and used this as a p-value to test the relative nestedness of submatrices. The same was repeated for the matrices of 6 and 5 columns. The submatrix with the lower probability is the most nested, and is assumed to contain the centre of origin.

f – After having identified the group of areas with the highest nestedness, we considered the area with the highest richness within this set as the centre of origin.

As a measure of nestedness we used the temperature metric T (Atmar & Patterson, 1993) because the weighed distance concept that is inherent to this metric seems particularly able to deal with large scale spatial data sets (Ulrich et al. 2009). T has recently been criticized for having unfavorable type I error probabilities (Almeida-Neto et al. 2008). However, in our study we do not refer to significance levels of single matrices, but to differences in Z-values between matrices. Thus, probability levels are of minor importance and do not outweigh the advantages of T. In addition, our approach is not affected by the influence of the size of the matrix under study, and the concrete eventuality of incurring type I or type II statistical errors consequent to the choice of the null model (e.g. Ulrich & Gotelli 2007), because both the matrix under study and the random matrices have the same sizes and are processed with the same algorithm.

In principle, meaningful major groups of areas within a larger set of areas can be recognized a priori by geographical, ecological, biogeographical or faunal boundaries, or a posteriori, by grouping areas with some statistical technique, such as ordination or classification methods. In our study, we used wide primary areas based on a wide range of organisms (including siganids), because they are considered most appropriate to reflect true biogeographical barriers (Bellwood & Wainwright 2002; Golokov et al. 1990). We are aware that we cannot test the biogeographical relevance of these primary areas, because they

Table I. – Matrix of presence/absence of Siganus species in the regions of Fig. 1.

| Species     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|-------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| argenteus   | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0  | 1  | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| canaliculatus | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| corallinus  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| dolius      | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| fuscescens  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| guttatus    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| javus       | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| labyrinthodes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| lineatus    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| luridus     | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| magnificus  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| niger       | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| puelloides  | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| puellas     | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| punctatissimus | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| punctatus   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| randalli    | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| rivulus     | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| sutor       | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| trispilos   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| unimaculatus | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| uspi        | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| vermiculatus | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| virgatus    | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| vulpinus    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| woodlandi   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

Total number of species 2 4 4 6 7 4 4 4 9 13 10 11 12 9 17 16 5 12 16 5 2 9 7 12
have been established also on the basis of siganid distributions (along with a wide array of other organisms), so that there is no complete statistical independence. However, the method can be applied to any geographical unit, at any scale.

We identified the following three groups of basic areas on a biogeographical basis: Western Indian ocean (areas 1-8), IMPA/east Asian coasts (areas 9-16), and Western and Central Pacific (areas 17-24) (Fig. 1). Such area groups were chosen taking into account the mid-Indian barrier, because of its crucial role in separating eastern and western marine fauna (see Golikov et al. 1990), and the major barriers separating the IMPA from the rest of the Pacific and from Australia (as summarized by Bellwood & Wainwright 2002).

To calculate Z-values, we used the software Nestedness (Ulrich 2006) with an equiprobable null model, 100 iterations to compute the standard deviation of the null model, a minimum distance of 0.5 to the borderline and matrix packed according to richness. Z-values below -2.0 indicate significant nestedness at the 5 % error level (Ulrich & Gotelli 2007).

Random selection of columns from the C-matrix was performed in the R environment using the “sample” function. Each set of 1000 random matrices was then processed with the Nestedness software (Ulrich & Gotelli 2007).

RESULTS

Table II reports the results of the nestedness analysis. Statistics resulting from the nestedness analysis performed on the three sets of 1000 random matrices are given in Table III. The p-values determined by comparing the null distribution of the Z-values with the values observed in the three groups of areas were as follows: 0.2 % in the case of the matrix corresponding to the region of IMPA and East Asian coasts; 1.7 % in the case of the matrix corresponding to the region of Western and Central Pacific; and 60.2 % in the case of the matrix corresponding to the region of Western Indian Ocean. Thus, the IMPA/East Asian coasts and the Western and Central Pacific are identified as the groups of primary areas that are significantly more nested than any other equal-sized random assortments of primary areas, whereas the Western Indian Ocean is an area of secondary colonization. The IMPA/East Asian coasts region is more nested than Western and Central Pacific, and it can be considered as the true centre of origin. In particular, area 15 (Philippines, the richest among those of the most nested group) is identified as the centre of origin, with a possible extension to area 16. By contrast, area 19, albeit very rich, is considered a secondary centre of diversity.

DISCUSSION

Multiple datasets agree in recognizing the IMPA as a centre of marine biodiversity for several taxa, with a primary peak of species richness in the central Philippines, and a secondary peak between peninsular Malaysia and Sumatra, thus suggesting that the IMPA is a major centre of speciation from which species dispersed to marginal locations (Bellwood & Hughes 2001, Mora et al. 2003, Carpenter & Springer 2005).

We demonstrated not only that the Philippines are the area with the largest number of species, but also that the species found in other areas are subset of the Philippine fauna. Woodland (1983, 1990, 1999) has already addressed the question of the biogeography of the Siganidae, distinguishing sister species that form parapatric pairs or triplets from the other species. He also distinguished species according to their degree of ecological dispersal.
Based on this idiographic approach (sensu Fattorini 2007) he came to the preliminary conclusion that the central Indo-Malaysian region was the evolutionary centre of origin of most siganid species. In contrast, the West Indian Ocean and Pacific islands were peripheral, secondary foci of speciation. Our analysis of the geographical distribution of siganids largely confirmed Woodland’s conclusions, substantiating the hypothesis of past speciation events in the western Indian Ocean, probably promoted by the effect of the mid-Indian barrier on siganid dispersion and colonization processes. However, among the species under study, only seven are present in both west and east of the mid-Indian barrier, while 17 are present exclusively at the east of the barrier. The lack of correspondence between the main phylogenetic split identified within the family and the pattern of species distribution east and west of the mid-Indian barrier suggests that the early radiation in Siganidae predates the origin of this disjunct distribution (Borsa et al. 2007, Kuriwa et al. 2007). By contrast, S. sutor, S. rivulatus and S. luridus, which occur only west of the barrier, could have originated allopatrically as a consequence of the isolative effect of the mid-Indian barrier.

If speciation events were mostly limited to a given area (the centre of origin), with successive colonization of peripheral areas, a highly nested structure would reasonably be expected in the global distribution of species. Alternatively, if speciation occurred in all areas demarcated by barriers, one would expect to observe a global distribution that was at most weakly nested. This would happen because each one of the metacommunities, after its isolation, would diverge time after time from its original condition as a perfect subsample of the original fauna. Our results are coherent with the first scenario depicted above because the assemblage of areas containing the centre of origin turned out to be more nested than any other area assemblage. The high relative nestedness observed in the assemblage of the Western and Central Pacific should therefore be explained by its closeness to the centre of origin and by the lack of a barrier as effective as western ones (i.e. the mid-Indian barrier) in separating it from IMPA.

ACKNOWLEDGEMENTS. – We are grateful to R Shannon for the linguistic revision, to M Pasi for help with the software, and to three anonymous reviewers for their comments on previous versions of this paper.

REFERENCES


Vie Milieu, 2011, 61 (2)


Received September 4, 2009
Accepted July 3, 2010
Associate Editor: T Changeux