ViBRANT e-Infrastructure

Work Package 8: Ecological and conservation data mobilisation

Milestone 8.24: Scientific Document with the results of the COMBER data

By: Christos Arvanitidis, Thanos Dailianis, Giorgos Chatzigeorgiou and Sarah Fawlwetter

INTRODUCTION

The most difficult problem in the contemporary biodiversity research is the collection of large datasets from various spatial and temporal scales and from multiple levels of the biological organization. It is widely accepted that most of the ecological information and data are collected in the framework of tem porally limited projects, simply because the collection costs are covered by the project funds. This explains why the series of datasets that are currently available are predominately discontinuous or unevenly spread, geographically, temporally or functionally. The latter is particularly obvious in the marine biodiversity data the collection costs of which are much higher than those of the terrestrial ones due to the diverse and expensive floating means as well as the specific sampling gears and methods employed. Several international projects targeting at continuous marine biodiversity data collection from specific habitats have been launched in the last couple of decades. The most widely known among those is the NaGISA project (National Geography in Shore Areas; http://www.nagisa.coml.org/) which operates under the umbrella of CoML (Census of Marine Life, http://www.coml.org/).

On the other hand, as the population of the professional taxonomists is diminishing, the mobilization of citizen scientists has become a key element to the success of the information and data collection process (e.g. Delaney et al. 2007, Hand 2010, Silvertown 2009, Trumbull et al. 2000).

However, the implementation of such citizen science projects in the marine environment currently faces two difficulties: (a) only the tidal zone can be approached by all citizens, and (b) the maximal depth safely reachable by recreational SCUBA divers is limited to 40 m. In the latter case, expensive diving equipment and certified training are required.

COMBER (Citizens' Network for the Observation of Marine BiodivERsity, working version at: http://www.comber.hcmr.gr; Alpha version at: http://vm4.her.hcmr.gr/comber/scratchpads-2.0/) is a pilot project that has been initiated under the ViBRANT e-infrastructure (FP7 project, 2010-2013) which attempts to support virtual research communities in biodiversity science. COMBER is now integrated into Scratchpads2 which is the platform used for development of all of ViBRANT products and is based on Drupal. COMBER aims at engaging citizen scientists in a coastal marine biodiversity observation network. The technical and basic characteristics of this pilot project have been recently published by Arvanitidis et al (2011).

The purpose of this document is to test the value of the collected datasets under COMBER by a single, well-defined scientific hypothesis: whether the species lists and their higher classifications reported by the citizen scientists are random samples of the broader lists of the region (Aegean, Eastern Mediterranean and Mediterranean).

MATERIALS & METHODS

The datasets used in the study are collected over the three years of the pilot project. The target groups are basically the SCUBA divers and snorkelers. As reported in the previous paper (Arvanitidis et al. 2011), fish species were chosen as a target taxon for the implementation of the pilot project since they are abundant and most frequently attract the attention and interest of the wide audience. However, the project can now include any taxon from any shallow marine habitat (<50m). The species observation and data collection was facilitated by usage of the commercial BIOWATCH underwater fish identification guide (http://www.bio-watch.com). The underwater fish card (Dounas 2009, Dounas and Koulouri 2011) includes the forty most common fish species of the Mediterranean coastal environment. Although COMBER was initially designed to be implemented on the Cretan coasts, the demand from users from all over Greece made it possible to expand its activity and collect datasets from a number of locations from the Aegean and Ionian Seas.

All analyses were performed on order abundance values in two scales (on average and summarised level according to the observation scale). In detail the raw date were contained order values 1-3, 4-10,

more than 10. Those order values corresponding to rare, common and abundant species populations respectively, for this those values were substituted by magnitude values of 10, thus values 1-3 substituted by 10, 4-10 substituted by 100 and more than ten substituted by 1000. Accordingly fourth root transformation and standardisation of the datasets were performed. Similarity matrices were calculated based on Jaccard's similarity coefficient. This coefficient is used to reveal biogeographic patterns (Legendre and Legendre 1998). Non-metric multidimensional scaling (nMDS) was then used to visualise the relationship between the datasets.

All analyses were performed on order abundance values in two scales (on average and summarised level according to the observation scale). In detail the raw date were contained order values 1-3, 4-10, more than 10. Those order values corresponding to rare, common and abundant species populations respectively, for this those values were substituted by magnitude values of 10, thus values 1-3 substituted by 10, 4-10 substituted by 100 and more than ten substituted by 1000. Accordingly fourth root transformation and standardisation of the datasets were performed. Similarity matrices were calculated based on Jaccard's similarity coefficient. This coefficient is used to reveal biogeographic patterns (Legendre and Legendre 1998). Non-metric multidimensional scaling (nMDS) was then used to visualise the relationship between the datasets.

All analyses were performed on abundance values ranked into levels: 0 (absent), 1-3 observed individuals, 4-10 individuals, more than 10 individuals. The last three ranked values, corresponding to rare, common and abundant species populations respectively, were substituted by orders of magnitude values. Therefore, the rare rank (1-3 observed individuals) was replaced by the actual value of 10, the rank of common (4-10) was substituted by 100 and the abundant one (>10) by 1000. Accordingly, fourth root transformation and standardisation of the datasets were performed before the matrices were eneter-ing to the analysis. Similarity matrices were calculated based on Jaccard's similarity coefficient. This coefficient is used to reveal biogeographic patterns (Legendre and Legendre 1998). Non-metric multi-dimensional scaling (nMDS) was then used to visualise the relationship between the datasets.
The primary hypothesis stated above was tested by means of the Taxonomic Distinctness indices (e.g. Warwick and Clark, 2001). We used 2 recently developed indices: (i) average taxonomic distinctness (AvTD, Δ^+), which is calculated by the formula $\Delta^+ = [\Sigma\Sigma i<j\ \omega ij] / [s(s-1)/2]$, where ωij is the phylogenetic/taxonomic path length between species i and j, and s is the number of species); (ii) variation in taxonomic distinctness (VarTD, Λ^+), defined as $\Lambda^+ = [\{\Sigma\Sigma i\neq j\ \omega ij2\} / \{s(s-1)/2\}] - (\Delta^+)^2$. By constructing a simulation distribution (funnel) from random subsets of fish species from the regional and sectoral, that is Mediterranean, eastern Mediterranean and Aegean inventories, both AvTD

and VarTD statistics, calculated from the fish species lists of the areas considered, can be tested for departure from expectation (e.g. Warwick & Clarke 2001). Values of AvTD and VarTD located within the 95% probability funnel indicate that fish species taxonomic/phylognetic diversity in the corresponding areas falls within the expected range and thus provides a way of testing for the criterion of randomness. These indices have been proved to be robust enough to perform both sample-size/sample-effort free diversity comparisons and to use the inventories in 'biological effects' monitoring studies in the future.

The various groups of the collected datasets were also compared by means of the non-metric multidimensional scaling technique. The significance of the compared groups was tested by means of the ANOSIM test, a variant of the non-parametric ANOVA (Clarke and Warwick, 1994).

RESULTS

A total of 409 COMBER activities, that is observations by either SCUBA diving or snorkeling, were registered in the system over the period May 2011 untill September 2013. Out of these observation activities, 75% were carried out by SCUBA diving and 25% by snorkeling. The majority of the 141 citizen scientists who took part to this pilot project come from Greece (49%), followed by divers from UK, Italy, Belgium and Germany, at percentages 8%, 8%, 9% and 5%, respectively. A 20% of the observation activities were performed by citizen scientists from other countries. The citizen scientists come from 15 countries, including countries from overseas such as the USA and Australia. The observers registered data show that the majority of these activities were performed just once (61%). Those which performed twice, 3-5, 6-10 or more than 10 times represent the 19%, 8%, 5% and 6% of the total observation activities, respectively. The data were collected by a total of 17 surveyed locations, with a total of 95 distinct observation sites. The total number of observations on fish species reached the number of 5,652 by the time the data were collected for the analysis.

In the following Figure, the frequencies of the most commonly observed species are demonstrated. The fish species most frequently observed by the citizen scientists were the *Coris julis*, *Chromis chromis* and *Thalassoma pavo*. They were reported by the participating citizen scientists more than 200 times, over the entire period of the pilot project. *Diplodus vulgaris*, *Sparisoma cretense* and *Oblada melanura* were observed more than 150 times.

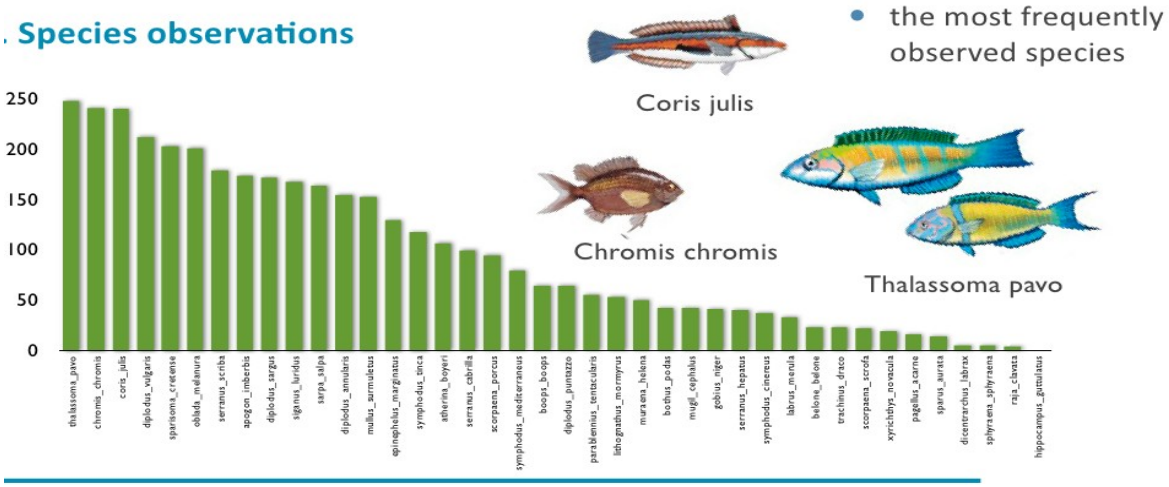


Figure 1. Species counts in all sites, over the entire period of the project: 2011-2013.

On the opposite site, the less frequently observed species were *Raja clavata*, *Sphyraena sphyraena* and *Hippocampus guttulatus*. The majority of the citizen scientists stated that they liked much both the project idea, as well as its implementation whereas only a 4% disliked the project's activities. Surprisingly, most of the volunteers had never participated in such projects in the past (64%) whereas a substantial number (28%) stated that they had been actively engaged in volunteers projects in the past. One of the important findings of the project's data on the volunteers' activities is that the reason for their participation to the COMBER project was their feeling of contribution to the monitoring of the marine biodiversity (58%). Another 20% of the volunteers stated they participated in order to gain new knowledge and experience while only 14% out of curiosity. Another interesting finding was that all of the volunteers found the process of the underwater fish identification relatively easy, a fact which largely facilitated the interpretation of the results.

As shown in the following Figure, most of the fish species lists fall under the lower 95% confidence limits set by the Mediterranean inventory, for the Average Taxonomic Distinctness expected distribution values.

However, the picture is reversed when the Variation in Taxonomic Distinctness is taken into account. Here, most of the calculated values fall inside the expected distribution, indicating that most of the fish species lists and their higher classification is a random sample of the respective Mediterranean one (Figure 2). Finally, Table 1 provides a summary of these findings.

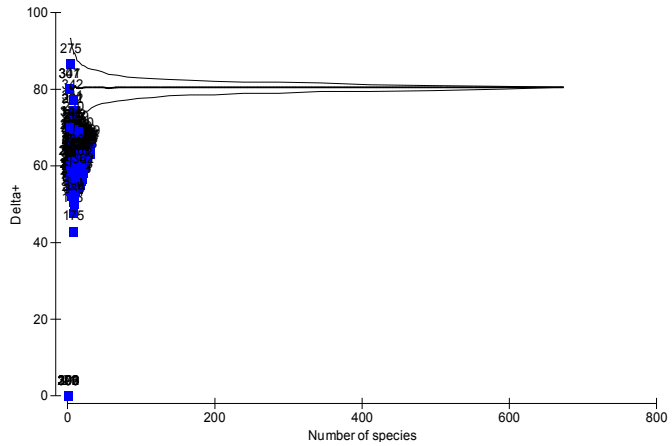


Figure 2. The 95% probability funnel for the average taxonomic distinctness (Δ^+) for all the fish species lists reported by the COMBER citizen scientists, as compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Mediterranean fish species and their higher phylogenetic/taxonomic classifications.

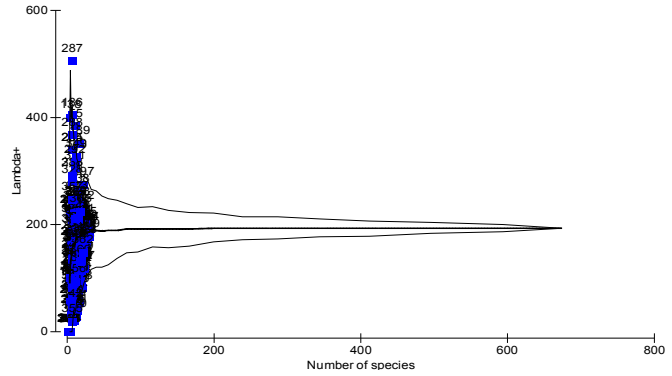


Figure 3. The 95% probability funnel for the variation in taxonomic distinctness (Λ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Mediterranean fish species and their higher phylogenetic/taxonomic classifications.

Table 1. Summary results of the Taxonomic Distinctness values calculated from the species lists reported by the COMBER participants, which fall or not within the expected distributions, as caclulated by the entire fish inventory of the Mediterranean Sea.

	In	**Out**
Delta	33	299
Lambda	326	36

The same trend in the values of both Taxonomic Distinctness indices was observed when the 95% confidence limits in the expected distributions was calculated by the fish inventory of the Eastern Mediterranean and of the Aegean Sea. This trend is provided in the following Figures and Tables.

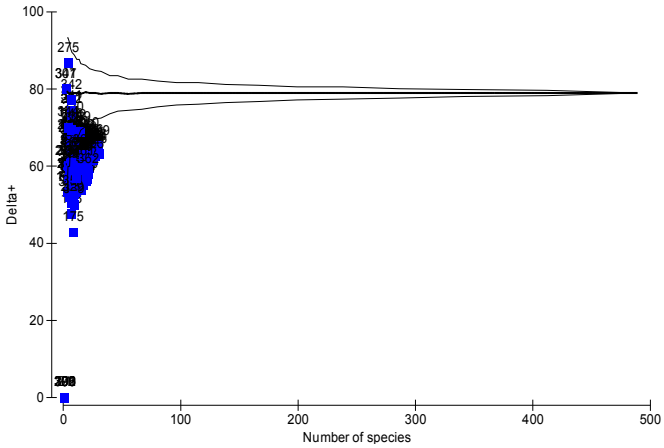


Figure 4. The 95% probability funnel for the average taxonomic distinctness (Δ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Mediterranean fish species inhabiting waters shallower than 50 m and their higher phylogenetic/taxonomic classifications.

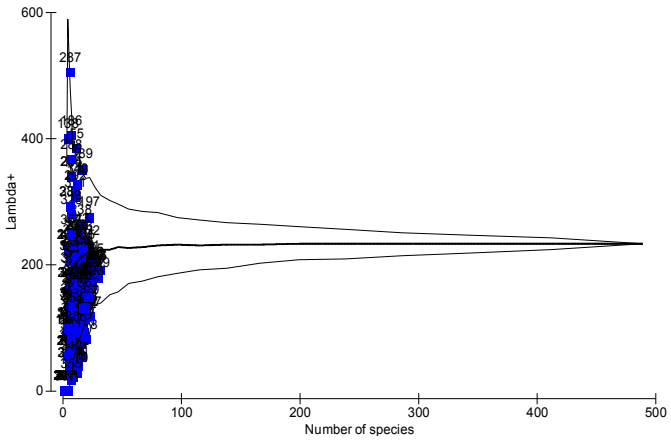


Figure 5. The 95% probability funnel for the variation in taxonomic distinctness (Λ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Mediterranean fish species inhabiting waters shallower than 50 m and their higher phylogenetic/taxonomic classifications.

Table 2. Summary results of the Taxonomic Distinctness values calculated from the species lists reported by the COMBER participants, which fall or not within the expected distributions, as caclulated by the Mediterranean fish species inhabiting waters shallower than 50 m.

	In	**Out**
Delta	58	304
Lambda	329	33

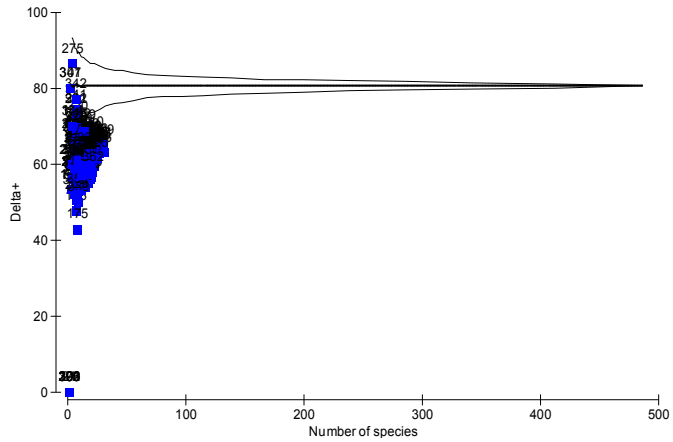


Figure 6. The 95% probability funnel for the average taxonomic distinctness (Δ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Aegean fish species and their higher phylogenetic/taxonomic classifications.

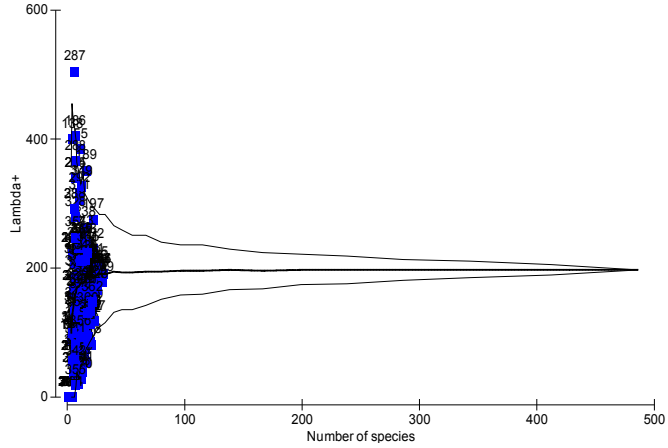


Figure 7. The 95% probability funnel for the variation in taxonomic distinctness (Λ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Aegean fish species and their higher phylogenetic/taxonomic classifications.

Table 4. Summary results of the Taxonomic Distinctness values calculated from the species lists reported by the COMBER participants, which fall or not within the expected distributions, as calculated by the entire fish inventory of the Aegean Sea.

	In	**Out**
Delta	34	328
Lambda	325	37

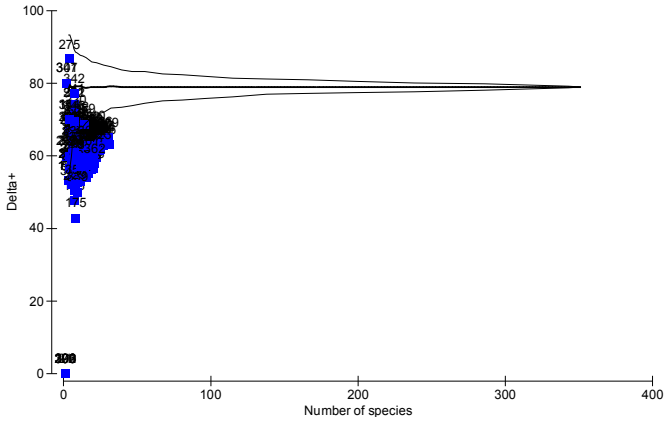


Figure 8. The 95% probability funnel for the average taxonomic distinctness (Δ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Aegean fish species inhabiting waters shallower than 50 m and their higher phylogenetic/taxonomic classifications.

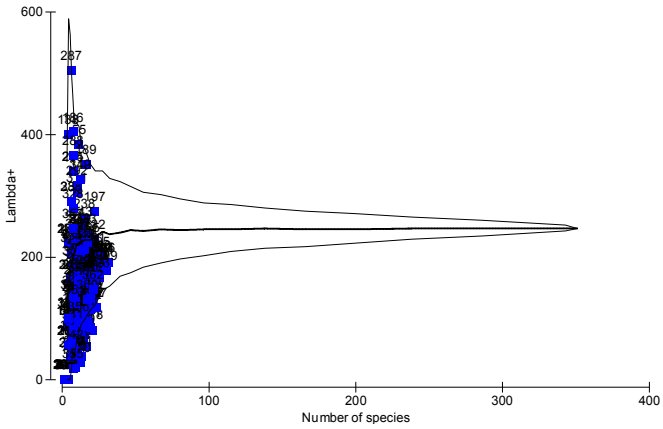


Figure 9. The 95% probability funnel for the variation in taxonomic distinctness (Λ^+) for all the fish species lists compiled from the data collected over the entire period of the COMBER project: 2011-2013. Funnel's confidence limits are calculated from random samples of the Aegean fish species inhabiting waters shallower than 50 m and their higher phylogenetic/taxonomic classifications.

Table 5. Summary results of the Taxonomic Distinctness values calculated from the species lists reported by the COMBER participants, which fall or not within the expected distributions, as calculated by the Aegean fish species inhabiting waters shallower than 50 m.

	In	**Out**
Delta	61	301
Lambda	324	38

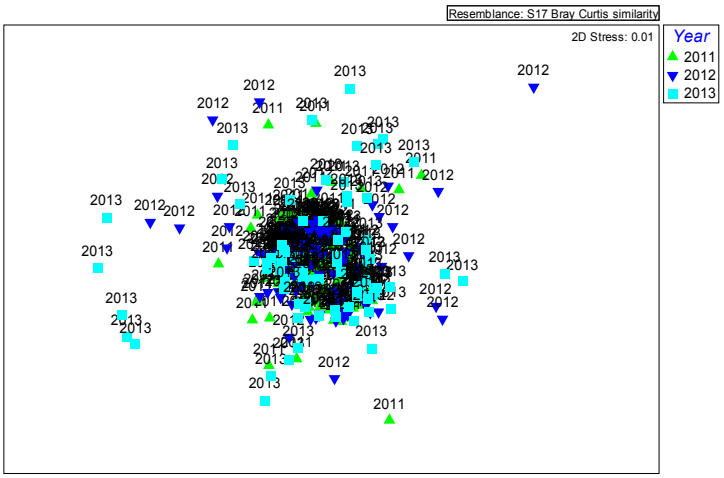


Figure 10. Non-metric multidimensional scaling (NMDS) ordination plot resulting from the resemblance matrix of the species composition in the samples collected by the divers and snorkelers over the entire period of the COMBER project: 2011-2013. Significant differences in the aggregated species lists collected in the three years are presented in the following Table.

Table 6. Results of the ANOSIM test on the significance of differences in observation between the years of the implementation of the COMBER citizen science project.

Global Test

Sample statistic (Global R): 0.115

Significance level of sample statistic: 0.1%

Number of permutations: 999 (Random sample from a large number)

Number of permuted statistics greater than or equal to Global R: 0

Pairwise Tests

Groups	R Statistic	Significance p-value	Possible Permutations	Actual Permutations	Number Observed
2011, 2012	0.034	0.01	Very large	999	0
2011, 2013	0.235	0.01	Very large	999	0
2012, 2013	0.12	0.01	Very large	999	0

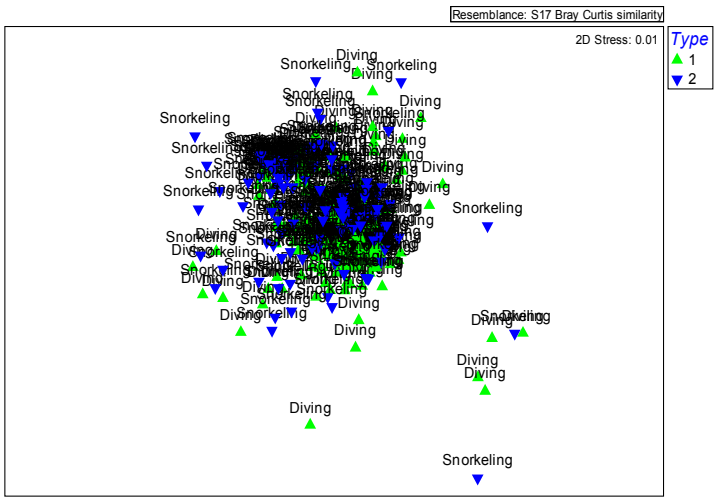


Figure 11. Non-metric multidimensional scaling (NMDS) ordination plot resulting from the resemblance matrix of the species composition in the samples collected by the divers and snorkelers over the entire period of the COMBER project: 2011-2013. Significant differences in the aggregated samples between the divers and snorkelers are presented in the following Table.

Table 7. Results of the ANOSIM test on the significance of differences in observation between the two groups of citizen science participating to COMBER activities: SCUBA divers and snorkelers, over the entire period of the project.

Global Test

Sample statistic (Global R): 0.186

p-value: 0.01

Number of permutations: 999 (Random sample from a large number)

Number of permuted statistics greater than or equal to Global R: 0

Table 8. Comparison of the Taxonomic Distinctness Indices values calculated by the data delivered by the two citizen science groups: SCUBA divers and snorkelers. In each cell, the first number refers to the samples located outside the lower 95% expected distribution confidence interval and the second number refers to the samples located within the confidence intervals.

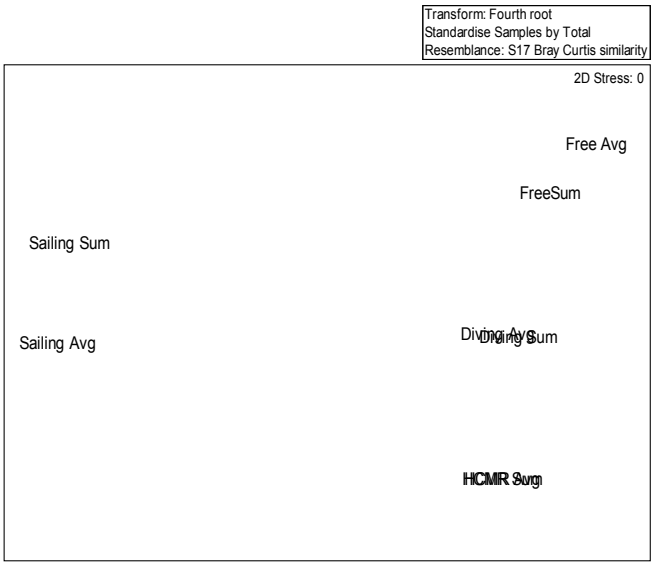
	Aegean 50m		**East Med 50m**		**Med 50m**	
	Delta	Lambda	Delta	Lambda	Delta	Lambda
Diving	248 / 12	9 / 251	248 / 12	7 / 253	249 / 11	7 / 253
Snorkelling	89 / 12	44 / 57	91 / 10	31 / 70	9 / 92	30 / 71

No patterns are observed between those values located outside the funnel limits and those located into the funnel. The only observable pattern can be found in the case of comparison between diving and snorkeling data where only for Lambda index the majority of the values located outside the lower funnel limit are coming from snorkeling.

Figure 12. Comparison of the datasets collected by the skilled divers and the most active snorkelers, who brought more than four datasets to the COMBER system, by means of a non-metric MDS.

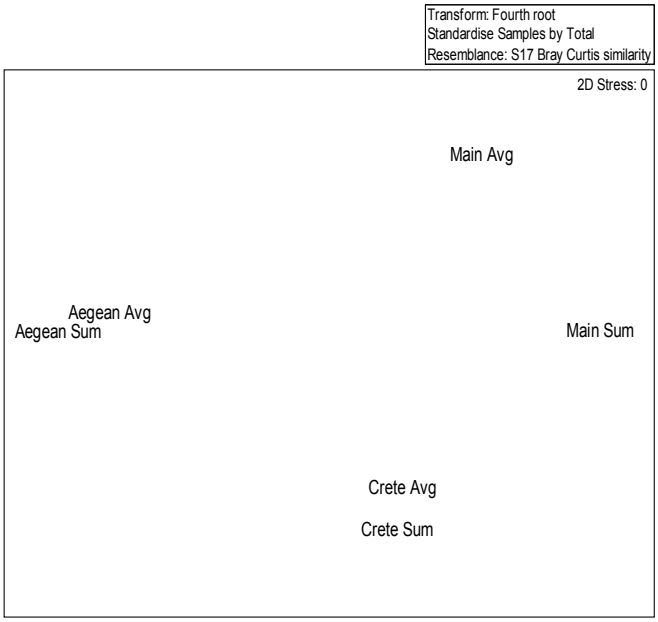
It is obvious that the results of the skilled citizen scientists divers are different from that of the snorkelers and that in the second case the datasets appear very diverse. The ANOSIM test confirmed the siginificance of these results (Global R = 1 , $p<0.01$).

Figure 13. Comparison of the datasets collected by the different groups of citizen scientists in relation to their origin, by means of a non-metric MDS: diving clubs, HCMR employees or invited scientists, sailing companies and free users, in two scales of data analysis: sums (all datasets pooled) and average (datasets collected by each category divided by the number of datasets).



The ANOSIM test confirmed the significance of these results (Global R = 1 , $p<0.01$).

Figure 14. Comparison of the datasets collected by the different locations in Greece, by means of a non-metric MDS: Crete, islands and sites in the proximity of the mainland groups, at two scales of data analysis (as in the previous Figure) .



The ANOSIM test confirmed the significance of these results (Global R = 1 , $p<0.01$).

DISCUSSION & CONCLUSIONS

After three years of the COMBER implementation, the datasets collected show the following trends:

1. COMBER as a pilot citizen science project attracted a satisfying number of volunteers who collected plentiful datasets not only from the Cretan coasts but from many locations all over the Aegean and Ionian Seas. COMBER provided volunteering citizen scientists with a user-friendly web application, based on Scratchpads2, which facilitates the collection and storage of their datasets as well as means for their communication as a community.
2. The application of the Taxonomic Distinctness Indices to the datasets collected by citizen scientists show that these datasets can be considered as reliable, up to a certain degree, excluding perhaps those of the snorkelers. This is the primary outcome of the Variation in Taxonomic Distinctness, which measures the variation in the species relatedness and not on the Average version of the Indices. This result certainly does not imply that the data collected by the snorkelers are faulty. Certainly, however, they are different and this can be attributed to the limitations regarding the depth at which observations with snorkeling can take place. Hence, snorkelers actually observe less fish species than scuba divers.
3. Differences also emerge when the datasets collected are compared by means of non-metric multidimensional scaling: again, the group of SCUBA divers differs from that of the snorkelers but differences can be also be observed between the years of the project's implementation, between the most skilled citizen-scientists, between the origin of the divers and between the main locations from which the datasets were collected.

The project established its existence and laid the foundations for its continuation in the future through a number of similar projects starting up early on next year.

LITERATURE CITED

Arvanitidis C, Faulwetter S, Chatzigeorgiou G, Penev L, Bánki O, Dailianis T, Pafilis E, Kouratoras M, Chatzinikolaou E, Fanini L, Vasileiadou A, Pavloudi C, Vavilis P, Koulouri P, Dounas C (2011) Engaging the broader community in biodiversity research: the concept of the COMBER pilot project for divers in ViBRANT. *Zookeys* 150: 211–229

Clarke KR, Warwick RM (1994) *Change in marine communities: an approach to statistical analysis and interpretation*. Natural Environment Research Council, Plymouth Marine Laboratory, Plymouth

Dailianis T, Faulwetter S, Chatzigeorgiou G, Plaitis W, Prentiss N, Dounas C, Koulouri P, Artemiou S, Vlasiou A, Pinsoglio G, Arvanitidis C (2013) Engaging citizen scientists for the assessment of the marine biodiversity: experience from the ongoing project. COMBER website: http://www.comber.hcmr.gr/?q=content/project-materials

Dounas C (2009) Illustrated guide for the identification of organisms in the field. *International patent, WO 2009/144516, World Intellectual Property Organisation*, International Bureau, 3 December 2009.

Dounas C, Koulouri P (2011) Mediterranean coastal fishes: an illustrated snorkeler's guide. *Kaleidoskope – BIOWATCH Editors*, Heraklion, Crete, Greece, 100+x pp (In Greek).

Faulwetter S, Chatzigeorgiou G, Galil BS, Arvanitidis C (2011) An account of the taxonomy

Warwick RM, Clarke KR (2001) Practical measures of marine biodiversity based on relatedness of species. Oceanography and Marine BIOLOGY: An Annual Review 39:207–231