




## RESOURCE ARTICLE

# SPART: A versatile and standardized data exchange format for species partition information

Aurélien Miralles<sup>1</sup>  | Jacques Ducasse<sup>2</sup> | Sophie Brouillet<sup>3</sup> | Tomas Flouri<sup>3</sup> |  
Tomochika Fujisawa<sup>4</sup> | Paschalia Kapli<sup>3</sup> | L. Lacey Knowles<sup>5</sup> | Sangeeta Kumari<sup>6</sup> |  
Alexandros Stamatakis<sup>7,8</sup> | Jeet Sukumaran<sup>9</sup> | Sarah Lutteropp<sup>7</sup> | Miguel Vences<sup>6</sup>  |  
Nicolas Puillandre<sup>1</sup> 

<sup>1</sup>Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

<sup>2</sup>Independent Researcher, Paris, France

<sup>3</sup>Department of Genetics, Evolution and Environment, Centre for Life's Origins and Evolution, University College London, London, UK

<sup>4</sup>Center for Data Science Education and Research, Shiga University, Shiga, Japan

<sup>5</sup>Department of Ecology and Evolution, University of Michigan, Ann Arbor, MI, USA

<sup>6</sup>Braunschweig University of Technology, Zoological Institute, Braunschweig, Germany

<sup>7</sup>Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

<sup>8</sup>Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

<sup>9</sup>Biology Department, LS 262, San Diego State University, San Diego, CA, USA

## Correspondence

Aurélien Miralles and Nicolas Puillandre, Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 75005 Paris, France.

Emails: miralles.skink@gmail.com and nicolaspuillandre@gmail.com

Miguel Vences, Braunschweig University of Technology, Zoological Institute, Braunschweig, Germany.

Email: m.vences@tu-braunschweig.de

## Funding information

H2020 European Research Council, Grant/Award Number: 865101; Deutsche Forschungsgemeinschaft, Grant/Award Number: VE247/20-1

## Abstract

A wide range of data types can be used to delimit species and various computer-based tools dedicated to this task are now available. Although these formalized approaches have significantly contributed to increase the objectivity of species delimitation (SD) under different assumptions, they are not routinely used by alpha-taxonomists. One obvious shortcoming is the lack of interoperability among the various independently developed SD programs. Given the frequent incongruences between species partitions inferred by different SD approaches, researchers applying these methods often seek to compare these alternative species partitions to evaluate the robustness of the species boundaries. This procedure is excessively time consuming at present, and the lack of a standard format for species partitions is a major obstacle. Here, we propose a standardized format, SPART, to enable compatibility between different SD tools exporting or importing partitions. This format reports the partitions and describes, for each of them, the assignment of individuals to the "inferred species". The syntax also allows support values to be optionally reported, as well as original trees and the full command lines used in the respective SD analyses. Two variants of this format are proposed, overall using the same terminology but presenting the data either optimized for human readability (matricial SPART) or in a format in which each partition forms a separate block (SPART.XML). ABGD, DELINEATE, GMYC, PTP and TR2 have already been adapted to output SPART files and a new version of LIMES has been developed to import, export, merge and split them.

## KEYWORDS

integrative taxonomy, LIMES v2.0, SPART, species delimitation programs, species partition format

## 1 | INTRODUCTION

Species delimitation (SD) is a burgeoning, fully fledged research field in systematic biology (Camargo & Sites, 2013; Ducasse et al., 2020; Flot, 2015; Sites & Marshall, 2003). SD benefits from the interpretation of species as independent evolutionary lineages (de Queiroz, 1998, 2007) that can be distinguished from each other using a variety of operational SD criteria (Samadi & Barberousse, 2006). In integrative taxonomy (Dayrat, 2005; Padial et al., 2010), various lines of evidence and a wide range of data types can be used in formalised analytical workflows to propose species hypotheses, from DNA barcodes to phylogenomic data, discrete morphological characters, morphometric measurements, ecological traits, geographic occurrence, bioacoustic signals, metabolomic profiles, and others (Miralles et al., 2020).

If many, and among them the earliest, formalised SD procedures are mostly carried out manually, for example, by comparing trees with the geographic occurrence of individuals, calculating correlations between geographic and genetic distances, assessing steepness of hybrid zones, or seeking for correlation between genetic distance and morphological characters (Derkarabetian & Hedin, 2014; Dufresnes et al., 2015; Flot et al., 2010; Good & Wake, 1992; Miralles & Vences, 2013; Puillandre, Modica, et al., 2012; Vieites et al., 2009; Weisrock et al., 2010; Wiens & Penkrot, 2002), a substantial number of computer-based tools has been developed to delimit species, often based on statistical criteria. These programs can analyse large data sets, with a strong focus on the use of sequence data (Table 1). These methods have significantly contributed to increase the objectivity, repeatability, and speed of species delimitation inferences under different mathematical models and assumptions (e.g., multispecies coalescent model, DNA barcode gap, haplotype fields of recombination, cf. Carstens et al., 2013; Flot et al., 2010; Knowles & Carstens, 2007; Leavitt et al., 2015; de Queiroz, 1998, 2007; Rannala, 2015; Yang & Rannala, 2010).

Although the number and importance of SD tools is likely to sharply increase in the immediate future, they are not yet routinely used in the majority of alpha-taxonomic studies that result in the naming of over 15,000 new species of organisms every year (Miralles et al., 2020). One obvious shortcoming is the lack of interoperability among the various independently developed SD programs, and the lack of comprehensive software suites that offer various user-friendly features, such as those for data visualization and comparison of results across methods. For instance, incongruent species partitions resulting from different SD approaches applied to a given data set are common. They can even be significant, if not striking in some cases (such as excessive splitting or lumping leading to highly different number of species delimited; Carstens et al., 2013; Miralles & Vences, 2013; Dellicour & Flot, 2015; Kapli et al., 2016; Postaire

et al., 2016; Renner et al., 2017 for empirical cases; Sukumaran & Knowles, 2017; Chan et al., 2020; Luo et al., 2018, Mason et al., 2020; Zhang et al., 2011 for more methodological studies on SD limitations) and may depend on the biological properties of the species (Ahrens et al., 2016; Eberle et al., 2019; Esselstyn et al., 2012; Fujisawa & Barraclough, 2013). Integrative taxonomists will seek to compare these alternative species partitions across SD approaches (but see Rannala, 2015), and eventually estimate their robustness by integrating other data sources (morphological variation, geographic distribution, etc), in order to make an informed choice – a procedure that is excessively time consuming at present, given the lack of a standard format for species partitions.

The main output of species delimitation, and therefore of any SD program, is a species partition. The term “partition” here follows the set theory concept: the organization of a set of elements into mutually-exclusive and jointly-comprehensive subsets, not including the empty subset (Hrbacek & Jech, 1999). In an SD application, the elements are individuals (i.e., samples or specimens), and a specific species delimitation hypothesis is a particular assignment (i.e., a partition) of all these individuals to different subsets. Each subset usually corresponds to a distinct inferred species, but categories resulting from an SD analysis have been referred to by various other terms, such as primary species hypothesis, operational taxonomic unit (OTUs), barcode index number (BINs; Ratnasingham & Hebert, 2013), or even cluster (without any particular status [Figure 1]), but all of them match the aforementioned definition of a subset.

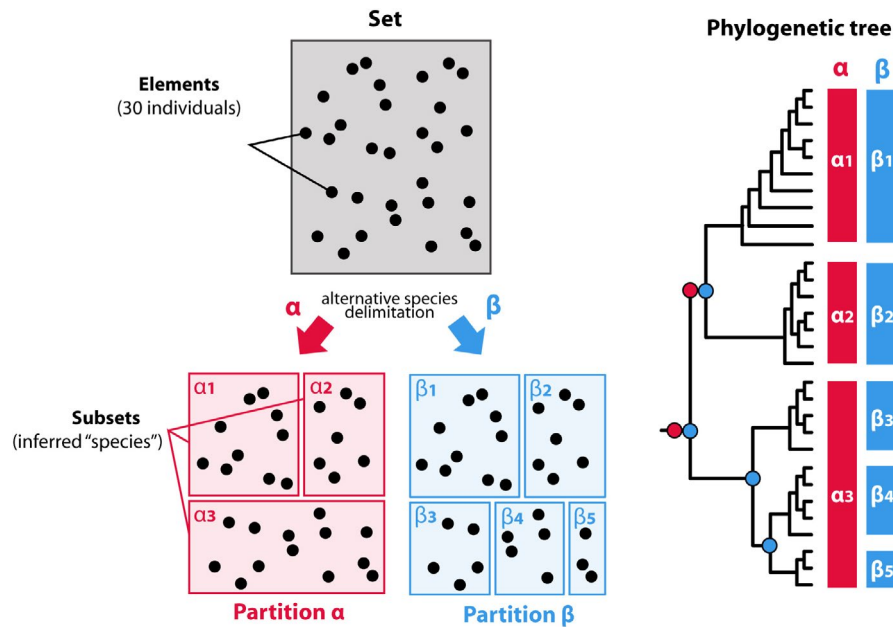
Furthermore, new tools producing *de novo* species partitions (i.e., directly aggregating individuals into species hypotheses) have recently been developed, and some of these, such as DELINEATE (Sukumaran et al., 2021) also statistically evaluate and compare the support of each possible species partition. Other methods statistically compare competing species hypotheses that have been defined a priori (primary species hypothesis testing), and these programs require a species partition as input. Some SD methods may assign scores, either to the entire inferred partition (e.g., ASAP-score in the program ASAP; Puillandre et al., 2021), to the distinctiveness of each subset from the others (e.g., posterior probabilities in the programs BPP and bPTP; Yang & Rannala, 2010; Zhang et al., 2013), or to the presence of each individual in a given subset (e.g., probability of placement in calculation of BINs, Ratnasingham & Hebert, 2013).

## 2 | A STANDARDIZED SPECIES PARTITION FORMAT (SPART)

Typically, each SD program exports the resulting species partitions in its own idiosyncratic format. Some, for instance, provide a table of assignments of individual specimens to the subsets (e.g., GMYC

**TABLE 1** Automated tools dedicated to species delimitation. Abbreviations used: mtDNA, mitochondrial DNA; nDNA, nuclear DNA. Note that for GMYC, PTP and DELINEATE, GUI-driven versions with SPART implementation have been prepared in the context of the iTaxoTools project but SPART output is not yet provided by all available versions. Other programs (ABGD, ASAP, TR2) already include native SPART output

Tools	General principle	Hypothetical partition needed as an input (a priori species assignment)	Optimal data sets and format	SPART implementation	References
GMYC (mGMYC and bGMYC)	General mixed Yule-coalescent model	No	mtDNA – ultrametric gene tree	Yes	Pons et al., (2006), Fontaneto et al., (2007), Monaghan et al., (2009)
BPP, iBPP	Multispecies coalescent model	Both options are possible	nDNA – multilocus alignments + (optionally in iBPP) matrix of morphological characters	In preparation	Yang and Rannala (2010, 2014), Solis-Lemus et al., (2015)
SPEDESTEM	Maximum likelihood and information theory	Yes	nDNA – ultrametric gene trees from multiple loci (nwk)	No	Ence and Carstens (2011)
ABGD	DNA barcode gap detection	No	mtDNA – sequence alignment or distance matrix	Yes	Puillandre, Lambert, et al., (2012)
SPECIES DELIMITATION	Coalescence/tree based approach	Yes	Topology (ultrametric tree)	No	Masters et al., (2011)
BINs	DNA barcode distance threshold +Markov clustering.	No	mtDNA – sequence alignment	No	Ratnasingham and Hebert (2013)
PTP (mPTP and bPTP)	Multivariate Poisson tree processes model	No	Nonultrametric tree (nwk or NEXUS tree)	Yes (mPTP and bPTP)	Zhang et al., (2013), Kapli et al., (2016)
DISSECT	Multispecies coalescent model	No	nDNA – multilocus alignments	No	Jones et al., (2014)
TR2	Multispecies coalescent model	No	nDNA – rooted gene trees from multiple loci (nwk)	Yes	Fujisawa et al., (2016)
STACEY	Multispecies coalescent model	No	nDNA – multilocus alignments	No	Jones (2017)
SODA	Quartet frequencies, based on coalescent model	No	Multiple gene tree topologies	No	Rabiee and Mirarab (2021)
HaplowebMaker/ CoMa	Mutual allelic exclusivity	No	nDNA – multilocus alignments	No	Spöri and Flot (2020)
ASAP	Distance-based partitions +coalescent-based scoring	No	mtDNA – sequence alignment or distance matrix	Yes	Puillandre et al., (2021)
DELINEATE	Multispecies coalescent model	Yes	Rooted ultrametric tree (nwk or NEXUS)	Yes	Sukumaran et al., (2021)



**FIGURE 1** In mathematics, a partition of a set is a grouping of its elements into non-empty subsets, in such a way that every element is included in exactly one such subset. The main output of a species delimitation inference therefore corresponds to a partition, independently of the theoretical context, the biological input data, or the algorithms/models used. In our example, a set of 30 specimens is split by two different methods into two alternative partitions  $\alpha$  and  $\beta$ , corresponding to three and five putative species (subsets), respectively. For the sake of clarity, these two alternative species partitions are represented as boxes reported next to each “species clade” in a phylogenetic tree, with hypothetical speciation events highlighted by circles via a corresponding colour. Note that not all SD methods rely on a tree topology, and may therefore delimit nonmonophyletic units (e.g., methods based on morphological or molecular divergence)

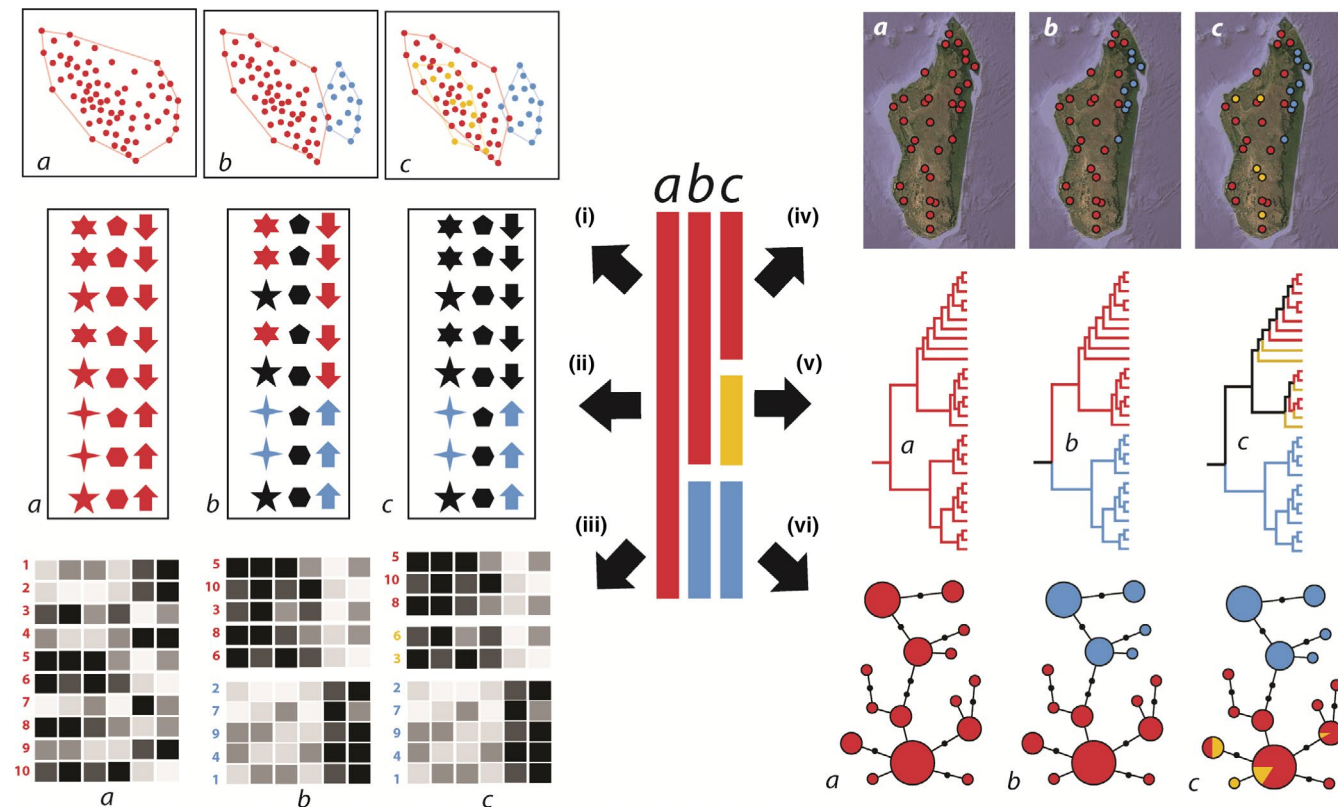
while others, conversely, list the different subsets with the included individuals (e.g., ABGD, PTP), whereas again others graphically report subsets on a tree topology (e.g., GMYC). These different formats may or may not include complementary data (e.g., scores, topologies, metadata, number of species delimited, etc.), and are not designed to be parsed by other tools for downstream analyses. Their manual conversion into a versatile and easily reusable plain text species partition (e.g., CSV) is not always straightforward. It can be particularly error prone and time consuming with large data sets, as species delimitations on several hundreds, or even thousands, of specimens are becoming common practice in molecular taxonomy (e.g., Ahrens et al., 2016; Christodoulou et al., 2020; García-Melo et al., 2019; Hoffmann et al., 2019; Renner et al., 2017; Sholihah et al., 2020).

We here propose a standardized species partition format, SPART, to enable compatibility between different tools producing (export) or using (import) species partitions. Our format facilitates:

1. Statistical comparison of different alternative species partitions such as their overall congruence, similarity or resolving power, identification of the subsets that are congruently delimited (currently implemented in the program LIMES v2.0; Ducasse et al., 2020);
2. Assessment of multiple competing SD hypotheses, including those used as input in e.g. BPP and DELINEATE to evaluate them (Sukumaran et al., 2021; Yang & Rannala, 2010);

3. Visualization and comparison of species partitions (e.g., DNA-based species partitions compared with manually-edited species partitions obtained from alternative methods and data such as principal component analysis of morphometry, haplotype networks, geographic distribution, habitat type, external phenetic similarity, or simply, current taxonomy);
4. Extraction, from original data files, of specific data for each subset under different species partition assumptions (e.g., lists of molecular and morphological diagnostic character states, descriptive statistics characterizing each of the inferred species, or ecological or distributional traits); and
5. Potential taxonomic reassignment of specimens in databases.

More generally, the SPART format is designed to be versatile and fully integrative in the sense that it can include any species partition descriptors, independently of the method or data-type used to generate the species partition (Figure 2). SPART does not convey any interpretation on the quality of the species partition, nor on the pros and cons of the methods used to define them, but is simply a common format that seeks homogenising the way species partitions are recorded. It can therefore be implemented in any method used to generate one or several species partitions as output. Likewise, any method using (analysing, comparing, automatically reassigning or graphically representing) multiple subsets of specimens might benefit from being able to import SPART files as input data.



**FIGURE 2** Illustration of exemplary potential applications of a species partition (SPART) file. If it can be parsed by other programs, SPART might facilitate the exploration of taxonomic data sets under various delimitation assumptions, such as (i) morphometric principal component analysis (PCA), (ii) automated extraction of diagnostic traits (three qualitative morphological characters with various states in this example), (iii) a heatmap of meristic morphological traits (for a visual exploration of the phenotypic variability), (iv) a distribution map, (v) a mitochondrial DNA-based phylogenetic tree, or (vi) a haplotype network reconstructed from nuclear-encoded DNA sequences. In the present example, among three putative alternative partitions (a, b and c), the partition b seems to represent the most plausible partition from a taxonomic perspective, as the distinctiveness of its two subsets is unambiguously supported by each of the six complementary approaches

### 3 | MATRICIAL AND SERIAL IMPLEMENTATION OF THE SPART FORMAT

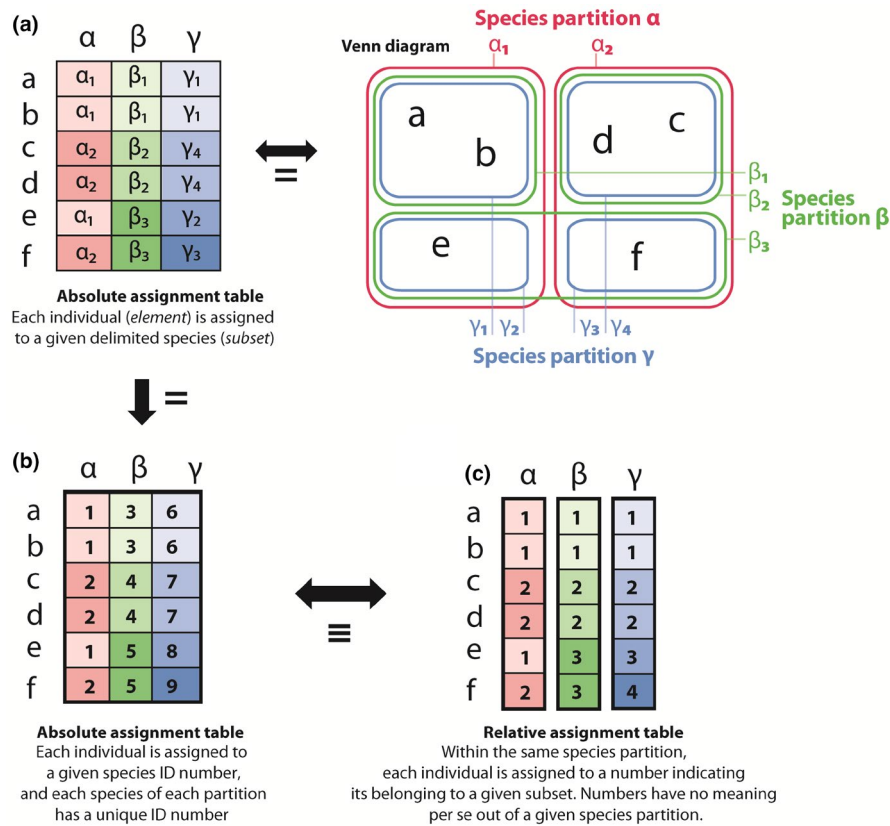
SPART files include information on one or multiple species partitions for a given set of elements (i.e., individuals) and use standardized terminology to denote the number of species partitions included in the file ("N\_partitions") and for each partition, the number of individuals ("N\_individuals"), number of subsets ("N\_subsets"), and the assignment of individuals to subsets ("Assignment") (Figure 3, Appendix S1). The syntax also allows to optionally include support values for species partitions, subsets, and the assignment of individuals to subsets, as well as original trees and the full command line used in the respective SD analyses, the program version number as well as comments and species partition comparison indices as calculated with LIMES 2.0, a new version of LIMES (Ducasse et al., 2020) recently published.

To account for the diversity of possible future applications, we propose two variants of the SPART format (for details see Appendix S1). Both of these use largely the same terminology but represent the data differently:

The first SPART variant is optimized for human readability and its syntax has been designed to be compatible with Nexus (a widely used data format in phylogenetic inference software: Maddison et al., 1997). This allows to include SPART specifications as blocks in Nexus files if required by future applications. If information from multiple partitions is included, then it is combined into a single block, presenting the respective assignments and assignment scores per individual from different species partitions concatenated on a single line, separated by separator symbols. This enables easy manual transformation into a spreadsheet format if required. Due to the presentation of information from multiple partitions in one block as a concatenated matrix, we denote this variant as matricial SPART format, or simply SPART.

The second SPART variant is optimized for machine readability, and relies on XML (eXtensible Markup Language), a lightweight data-interchange format that can be easily parsed and written by software tools, while it can still be read and written by humans as well. When information from multiple partitions is included, each partition forms a separate block containing information on the number of subsets, individual assignments and assignment scores. We therefore denote this variant as SPART.XML format.





**FIGURE 3** The SPART format can combine alternative species partitions of a same set of individuals (elements) into a unique multiple species partition file. (a) Example of set comprising six individuals split by three distinct SD analyses, resulting in three distinct species partitions ( $\alpha$ ,  $\beta$  and  $\gamma$ ). All these species partitions are hierarchically compatible (i.e., they conform to the mathematical definition of nested sets), with the exception of the pair  $\alpha$  -  $\beta$  (Venn diagram representing the alternative species partitions on the right, and corresponding assignment table on the left). These alternative species partitions can be coded in SPART either (b) by using a unique numbering for all the three species partitions (so that each species partition has its own set of species (subset) numbers) or (c) by using one numbering system per species partition. The latter representation allows combining different species partitions into a multiple species partition file without having to adjust each species or cluster number (subset). Both (b) and (c) are fully equivalent in SPART format, because the coding of each partition is independent from the others (subset assignment numbers have no meaning per se, they only indicate, within each partition, the common assignment to a specific subset)

## 4 | TOOLS ALREADY IMPLEMENTING SPART AND FUTURE PERSPECTIVES

The proposed format is already implemented in several widely-used SD programs. The matricial SPART output file is already generated by GUI-driven standalone versions (<https://github.com/iTaxoTools>; <http://itaxotools.org/>) of ABGD, ASAP, GMYC, PTP, mPTP, TR2 and DELINEATE (Vences et al., 2021), by the native Python version of TR2, and in the web versions of ABGD and ASAP; and in progress for the Python versions of GMYC and PTP. The implementation of the SPART.XML output will become available by the end of 2021 for ABGD and ASAP. Furthermore, the species partition comparison tool LIMES v2.0 has been expanded to import, export and convert matricial SPART files (SPART.XML files will be implemented by the end of 2021), in particular to (1) compare, by calculating indices (e.g., Ctax, Ratx, Match Ratio, cf. Ducasse et al., 2020) for species partitions from SPART files (including each one or several species partitions); (2) merge species

partitions included in different SPART files into one SPART file, (3) import species partition(s) table(s) from spreadsheet editors such as Microsoft EXCEL and save it (them) into a single SPART file. A new software tool named SPARTMAPPER has also been developed (Vences et al., 2021); it takes SPART files as input along with a tab-delimited series of geographical coordinates linked to specimen names, plots the distribution of alternative delimited species on a map, and exports a.kml file to visualize this information in Google Earth.

In the context of future work, we envisage the development of visualization tools to automatically illustrate information from species partitions along with support values and phylogenetic hypotheses (Figure 1). There is still a long way to go before programs will be able to infer species based on combining evidence using different data sources such as genetics, morphology, ecology, behaviour, geographic distribution, etc. However, eventually, reliable computer-based, species delimitation procedures that mirror the procedures of integrative taxonomy will be at the core of next generation taxonomy

(Vences, 2020). Our SPART data exchange format would thus contribute to this next generation taxonomy, by simplifying computational approaches to completing the inventory of life on Earth.

## ACKNOWLEDGEMENTS

We are grateful to Susanne Renner who stimulated this work by leading the priority program SPP 1991 "Taxon-Omics" of the Deutsche Forschungsgemeinschaft (DFG), specifically in the context of a grant on taxonomic data integration and management (RE 603/29-1), and to many members of the Taxon-Omics consortium and Guillaume Achaz (MNHN) for fruitful discussion. MV and SK were supported by DFG grant VE247/20-1, NP by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 865101), and AS and SL by the Klaus-Tschira foundation.

## DATA AVAILABILITY STATEMENT

All new versions of the above-mentioned software implementing the SPART format are already available on Github (<https://github.com/iTaxoTools>) and further information is available on the iTaxo-tools website (<http://itaxotools.org>).

## ORCID

Aurélien Miralles  <https://orcid.org/0000-0002-2538-7710>

Miguel Vences  <https://orcid.org/0000-0003-0747-0817>

Nicolas Puillandre  <https://orcid.org/0000-0002-9797-0892>

## REFERENCES

- Ahrens, D., Fujisawa, T., Krammer, H. J., Eberle, J., Fabrizi, S., & Vogler, A. P. (2016). Rarity and incomplete sampling in DNA-based species delimitation. *Systematic Biology*, *65*, 478–494. <https://doi.org/10.1093/sysbio/syw002>
- Camargo, A., & Sites, J. Jr. (2013). Species delimitation: a decade after the renaissance. In: I. Y. Pavlinov (Ed.), *The species problem - ongoing issues* (pp. 225–247). InTech – Open Access publisher. ISBN: 978-953-51-0957. Available from <http://www.inte-chopen.com/books/the-species-problem-ongoing-issues/spe-cies-delimitation-a-decade-after-the-renaissance>
- Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, *22*, 4369–4383. <https://doi.org/10.1111/mec.12413>
- Chan, K. O., Hutter, C. R., Wood, P. L. Jr, Grismer, L. L., Das, I., & Brown, R. M. (2020). Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex. *Molecular Ecology*, *29*(20), 3970–3987. <https://doi.org/10.1111/mec.15603>
- Christodoulou, M., O'Hara, T., Hugall, A. F., Khodami, S., Rodrigues, C. F., Hilario, A., Vink, A., & Martinez Arbizu, P. (2020). Unexpected high abyssal ophiuroid diversity in polymetallic nodule fields of the northeast Pacific Ocean and implications for conservation. *Biogeosciences*, *17*, 1845–1876. <https://doi.org/10.5194/bg-17-1845-2020>
- Dayrat, B. (2005). Toward integrative taxonomy. *Biological Journal of the Linnean Society*, *85*, 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In D. J. Howard, & S. H. Berlocher (Eds.), *Endless forms: species and speciation* (pp. 57–75). Oxford University Press.
- de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*, 879–886. <https://doi.org/10.1080/10635150701701083>
- Dellicour, S., & Flot, J.-F. (2015). Delimiting species-poor data sets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. *Systematic Biology*, *64*, 900–908. <https://doi.org/10.1093/sysbio/syu130>
- Derkarabetian, S., & Hedin, M. (2014). Integrative taxonomy and species delimitation in harvestmen: a revision of the western North American genus *Sclerobunus* (Opiliones: Laniatores: Travunioidea). *PLoS One*, *9*, e104982. <https://doi.org/10.1371/journal.pone.0104982>
- Ducasse, J., Ung, V., Lecointre, G., & Miralles, A. (2020). LIMES: a tool for comparing species partition. *Bioinformatics*, *2282*–2283. <https://doi.org/10.1093/bioinformatics/btz911>
- Dufresnes, C., Brelsford, A., Crnobrnja-Isailović, J., Tzankov, N., Lymberakis, P., & Perrin, N. (2015). Timeframe of speciation inferred from secondary contact zones in the European tree frog radiation (*Hyla arborea* group). *BMC Evolutionary Biology*, *15*, 1–8. <https://doi.org/10.1186/s12862-015-0385-2>
- Eberle, J., Bazzato, E., Fabrizi, S., Rossini, M., Colomba, M., Cillo, D., Uliana, M., Sparacio, I., Sabatinelli, G., Warnock, R. C. M., Carpaneto, G., & Ahrens, D. (2019). Sex-biased dispersal obscures species boundaries in integrative species delimitation approaches. *Systematic Biology*, *68*(3), 441–459. <https://doi.org/10.1093/sysbio/syy072>
- Ence, D. D., & Carstens, B. C. (2011). SpedeSTEM: A rapid and accurate method for species delimitation. *Molecular Ecology Resources*, *11*, 473–480. <https://doi.org/10.1111/j.1755-0998.2010.02947.x>
- Esselstyn, J. A., Evans, B. J., Sedlock, J. L., Anwarali Khan, F. A., & Heaney, L. R. (2012). Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1743), 3678–3686.
- Flot, J.-F. (2015). Species delimitation's coming of age. *Systematic Biology*, *64*(6), 897–899. <https://doi.org/10.1093/sysbio/syv071>
- Flot, J.-F., Couloux, A., & Tillier, S. (2010). Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. *BMC Evolutionary Biology*, *10*, 372. <https://doi.org/10.1186/1471-2148-10-372>
- Fontaneto, D., Herniou, E., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., & Barraclough, T. G. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biology*, *5*, e87. <https://doi.org/10.1371/journal.pbio.0050087>
- Fujisawa, T., Aswad, A., & Barraclough, T. G. (2016). A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology*, *65*(5), 759–771. <https://doi.org/10.1093/sysbio/syw028>
- Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), 707–724. <https://doi.org/10.1093/sysbio/syt033>
- García-Melo, J. E., Oliveira, C., Da Costa Silva, G. J., Ochoa-Orrero, L. E., Garcia Pereira, L. H., & Maldonado-Ocampo, J. A. (2019). Species delimitation of Neotropical characins (Stevardiinae): Implications for taxonomy of complex groups. *PLoS One*, *14*(6), e0216786. <https://doi.org/10.1371/journal.pone.0216786>
- Good, D. A., & Wake, D. B. (1992). Geographic variation and speciation in the torrent salamanders of the genus *Rhyacotriton* (Caudata: Rhyacotritonidae). *University of California Publications in Zoology*, *126*, 1–91.
- Hofmann, E. P., Nicholson, K. E., Luque-Montes, I. R., Köhler, G., Cerrato-Mendoza, C. A., Medina-Flores, M., Wilson, L. D., & Townsend, J. H. (2019). Cryptic diversity, but to what extent? Discordance between

- single-locus species delimitation methods within mainland anoles (Squamata: Dactyloidae) of Northern Central America. *Frontiers in Genetics*, 10, 11. <https://doi.org/10.3389/fgene.2019.00011>
- Hrbacek, K., & Jech, T. (1999). *Introduction to set theory, third edition, revised and expanded. Monographs and Textbooks in pure and applied mathematics*, Vol. 220. Marcel Dekker Inc.
- Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology*, 74, 447. <https://doi.org/10.1007/s00285-016-1034-0>
- Jones, G., Aydin, Z., & Oxelman, B. (2014). DISSECT: An assignment free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, 31, 991–998. <https://doi.org/10.1093/bioinformatics/btu770>
- Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2016). Multi-rate Poisson Tree Processes for single-locus species delimitation under Maximum Likelihood and Markov Chain Monte Carlo. *Bioinformatics*, 33, 1630–1638. <https://doi.org/10.1093/bioinformatics/btx025>
- Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), 887–895. <https://doi.org/10.1080/10635150701701091>
- Leavitt, S. D., Moreau, C. S., & Lumbsch, H. T. (2015). The dynamic discipline of species delimitation: Progress toward effectively recognizing species boundaries in natural populations. In: D. Upreti, P. Divakar, V. Shukla, & R. Bajpai (Eds.), *Recent Advances in Lichenology* (pp. 11–44). New Delhi: Springer. [https://doi.org/10.1007/978-81-322-2235-4\\_2](https://doi.org/10.1007/978-81-322-2235-4_2)
- Luo, A., Ling, C., Ho, S. Y. W., & Zhu, C.-D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, 67(5), 830–846. <https://doi.org/10.1093/sysbio/syy011>
- Maddison, D. R., Swofford, D. L., & Maddison, W. P. (1997). NEXUS: An extensible file format for systematic information. *Systematic Biology*, 46, 590–621. <https://doi.org/10.1093/sysbio/46.4.590>
- Mason, N. A., Fletcher, N. K., Gill, B. A., Funk, C., & Zamudio, K. R. (2020). Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, 18(3), 269–280. <https://doi.org/10.1080/14772000.2020.1730475>
- Masters, B. C., Fan, V., & Ross, H. A. (2011). Species Delimitation - a Geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources*, 11, 154–157. <https://doi.org/10.1111/j.1755-0998.2010.02896.x>
- Miralles, A., Bruy, T., Wolcott, K., Scherz, M. D., Begerow, D., Beszteri, B., Bonkowski, M., Felden, J., Gemeinholzer, B., Glaw, F., Glöckner, F. O., Hawlitschek, O., Kostadinov, I., Nattkemper, T. W., Printzen, C., Renz, J., Rybalka, N., Stadler, M., Weibulat, T., ... Vences, M. (2020). Repositories for taxonomic data: where we are and what is missing. *Systematic Biology*, 69, 1231–1253. <https://doi.org/10.1093/sysbio/syaa026>
- Miralles, A., & Vences, M. (2013). New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PlosONE*, 8, e68242. <https://doi.org/10.1371/journal.pone.0068242>
- Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J., Lees, D. C., Ranaivosolo, R., Eggleton, P., Barraclough, T. G., & Vogler, A. P. (2009). Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, 58, 298–311. <https://doi.org/10.1093/sysbio/syp027>
- Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16. <https://doi.org/10.1186/1742-9994-7-16>
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55, 595–609. <https://doi.org/10.1080/10635150600852011>
- Postaire, B., Magalon, H., Bourmaud, C. A., & Bruggemann, J. H. (2016). Molecular species delimitation methods and population genetics data reveal extensive lineage diversity and cryptic species in Aglaopheniidae (Hydrozoa). *Molecular Phylogenetics and Evolution*, 105, 36–49. <https://doi.org/10.1016/j.ympev.2016.08.013>
- Puillandre, N., Brouillet, S., & Achaz, G. (2021). ASAP: Assemble species by automatic partitioning. *Molecular Ecology Resources*, 21(2), 609–620. <https://doi.org/10.1111/1755-0998.13281>
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21, 1864–1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x>
- Puillandre, N., Modica, M. C., Zhang, Y., Sirovich, L., Boisselier, M. C., Cruaud, C., Holford, M., & Samadi, S. (2012). Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology*, 11, 2671–3691. <https://doi.org/10.1111/j.1365-294X.2012.05559.x>
- Rabiee, M., & Mirarab, S. (2021). SODA: Multi-locus species delimitation using quartet frequencies. *Bioinformatics*, 36(24), 5623–5631. <https://doi.org/10.1093/bioinformatics/btaa1010>
- Rannala, B. (2015). The art and science of species delimitation. *Current Zoology*, 61, 846–853. <https://doi.org/10.1093/czoolo/61.5.846>
- Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS One*, 8, e66213. <https://doi.org/10.1371/journal.pone.0066213>
- Renner, M. A., Heslewood, M. M., Patzak, S. D., Schäfer-Verwimp, A., & Heinrichs, J. (2017). By how much do we underestimate species diversity of liverworts using morphological evidence? An example from Australasian *Plagiochila* (Plagiochilaceae: Jungermanniopsida). *Molecular Phylogenetics and Evolution*, 107, 576–593. <https://doi.org/10.1016/j.ympev.2016.12.018>
- Samadi, S., & Barberousse, A. (2006). The tree, the network, and the species. *Biological Journal of the Linnean Society*, 89(3), 509–521. <https://doi.org/10.1111/j.1095-8312.2006.00689.x>
- Sholihah, A., Delrieu-Trottin, E., Sukmono, T., Dahruddin, H., Risdawati, R., Elvira, R., Wibowo, A., Kusno, K., Busson, F., Sauri, S., Nurhaman, U., Zein, M. S. A., Fitriana, Y., Utama, I., Muchlisin, Z. A., Agnèse, J. F., Hanner, R., Wowor, D., Steinke, D., ... Hubert, N. (2020). Disentangling the taxonomy of the subfamily Rasborinae (Cypriniformes, Danionidae) in Sundaland through DNA barcodes. *Scientific Reports*, 10, 2818. <https://doi.org/10.1038/s41598-020-59544-9>
- Sites, J. W., & Marshall, J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, 18, 462–470. [https://doi.org/10.1016/S0169-5347\(03\)00184-8](https://doi.org/10.1016/S0169-5347(03)00184-8)
- Solís-Lemus, C., Knowles, L. L., & Ané, C. (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69, 492–507. <https://doi.org/10.1111/evo.12582>
- Spöri, Y., & Flot, J.-F. (2020). HaplowebMaker and CoMa: two web tools to delimit species using haplowebs and conspecificity matrices. *Methods in Ecology and Evolution*, 11(11), 1434–1438. <https://doi.org/10.1111/2041-210X.13454>
- Sukumaran, J., Holder, T. M., & Knowles, L. L. (2021). Incorporating the speciation process into species delimitation. *PLoS Computational Biology*, 17(5), e1008924. <https://doi.org/10.1371/journal.pcbi.1008924>
- Sukumaran, J., & Knowles, L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences USA*, 114(7), 1607–1612. <https://doi.org/10.1073/pnas.1607921114>
- Vences, M. (2020). The Promise of next-generation Taxonomy. *Megatata*, 1, 35–38. <https://doi.org/10.11646/megatata.1.1.6>
- Vences, M., Miralles, A., Brouillet, S., Ducasse, J., Fedosov, A., Kharchev, V., Kumari, S., Patmanidis, S., Puillandre, N., Scherz, M. D.,



- Kostadinov, I., & Renner, S. S. (2021). iTaxoTools 0.1: Kickstarting a specimen-based software toolkit for taxonomists. *Megataxa*, 6(2), 77–92. <https://doi.org/10.11646/megataxa.6.2.1>
- Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences*, 106(20), 8267–8272. <https://doi.org/10.1073/pnas.0810821106>
- Weisrock, D. W., Rasoloarison, R. M., Fiorentino, I., Ralison, J. M., Goodman, S. M., Kappeler, P. M., & Yoder, A. D. (2010). Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS One*, 5, e9883. <https://doi.org/10.1371/journal.pone.0009883>
- Wiens, J. J., & Penkrot, T. A. (2002). Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Systematic Biology*, 51, 69–91. <https://doi.org/10.1080/106351502753475880>
- Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107(20), 9264–9269. <https://doi.org/10.1073/pnas.0913022107>
- Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31, 3125–3135. <https://doi.org/10.1093/molbev/msu279>
- Zhang, C., Zhang, D. X., Zhu, T., & Yang, Z. (2011). Evaluation of Bayesian coalescent method of species delimitation. *Systematic Biology*, 60(6), 747–761. <https://doi.org/10.1093/sysbio/syr071>
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869–2876. <https://doi.org/10.1093/bioinformatics/btt499>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Miralles, A., Ducasse, J., Brouillet, S., Flouri, T., Fujisawa, T., Kapli, P., Knowles, L. L., Kumari, S., Stamatakis, A., Sukumaran, J., Lutteropp, S., Vences, M., & Puillandre, N. (2022). SPART: A versatile and standardized data exchange format for species partition information. *Molecular Ecology Resources*, 22, 430–438. <https://doi.org/10.1111/1755-0998.13470>