DR. AURÉLIEN MIRALLES (Orcid ID : 0000-0002-2538-7710)

DR. NICOLAS PUILLANDRE (Orcid ID : 0000-0002-9797-0892)

Article type : Resource Article

Journal : Molecular Ecology Resources

Subject area: Computer Programs

SPART, a versatile and standardized data exchange format for species partition information

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

[1]Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, CP 50, 75005 Paris, France

[2]Independent researcher, 49 rue Eugène Carrière, 75018 Paris, France

[3]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

[4]Center for Data Science Education and Research, Shiga University, 1-1-1 Banba, Hikone, 522-8522, Shiga, Japan

[5]Department of Ecology and Evolution, University of Michigan, Ann Arbor, MI 48109, USA

[6]Braunschweig University of Technology, Zoological Institute, Mendelssohnstraße 4, 38106 Braunschweig, Germany

[7]Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg Germany

[8]Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Am Fasanengarten 5, 76131 Karlsruhe, Germany

[9]Biology Department, LS 262, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4614, USA

***Corresponding authors**

Contact: miralles.skink@gmail.com, m.vences@tu-braunschweig.de, nicolaspuillandre@gmail.com

Short running title: SPART, a standardized format for species partition

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 to optionally report support values, 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.

Key words. Species delimitation programs; SPART; Species partition format; Integrative taxonomy ; LIMES v2.0

Introduction

Species delimitation (SD) is a burgeoning, fully fledged research field in systematic biology (Sites & Marshall 2003; Camargo & Sites 2013; Flot 2015, Ducasse et al. 2020). 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, e.g. 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 (Good & Wake 1992, Wiens & Penkrot 2002, Vieites et al. 2009, Flot et al. 2010, Weisrock et al. 2010, Puillandre et al. 2012a, Miralles & Vences 2013, Derkarabetian & Hedin 2014, Dufresnes et al. 2015), a substantial number of computer-based tools has been developed to delimit species, often based on statistical criteria. These programs can analyse large datasets, 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. de Queiroz 1998, 2007, Knowles & Carstens 2007, Yang & Rannala 2010, Flot et al. 2010, Carstens et al. 2013, Leavitt et al. 2015, Rannala 2015).

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 dataset 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; and Sukumaran & Knowles 2017, Chan et al 2020, Luo et al. 2018, Mason et al. 2020 and Zhang et al. 2011 for more methodological studies on SD limitations) and may depend on the biological properties of the species (Esselstyn et al. 2012, Fujisawa & Barraclough 2013; Ahrens et al. 2016 ; Eberle et al. 2019). 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, where *each subset corresponds to a distinct inferred species*. Categories resulting from an SD analysis have been referred to by various terms, such as primary species hypothesis, operational taxonomic unit (OTUs), barcode index number (BINs; Ratnasingham & Hebert 2013), or even cluster (without any particular status (Fig. 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).

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) 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 datasets, as species delimitations on several hundreds, or even thousands, of specimens are becoming common practice in molecular taxonomy (e.g., Ahrens et al. 2016, Renner et al 2017, Garcià-Melo et al. 2019, Hoffmann et al. 2019, Solihah et al. 2020, Christodoulou 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 (Yang & Rannala 2010, Sukumaran et al. 2020);

(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 (Fig. 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 at 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.

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_spartitions”) and for each partition, the number of individuals ("N_individuals"), number of subsets (“N_subsets”), and the assignment of individuals to subsets (“Assignment”) (Fig. 3, Supporting information 1). 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 Supporting information 1). 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.

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. in press), 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. in press); 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 (Fig. 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.

Acknowledgments

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.

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. doi:10.1093/sysbio/syw002

Camargo, A., & Sites, J. Jr. (2013). Species delimitation: a decade after the renaissance. In: The Species Problem - Ongoing Issues (ed. I. Y. Pavlinov). IntechOpen.

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecolology, 22,* 4369–4383. doi: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. doi: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. doi:10.5194/bg-17-1845-2020

Dayrat, B. (2005). Toward integrative taxonomy. *Biological Journal of the Linnean Society,* 85, 407–415. doi:10.1111/j.1095-8312.2005.00503.x

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. doi:10.1093/sysbio/syu130

de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In: D.J. Howard & S.H. Berlocher, S.H. (Eds.), *Endless Forms: Species and Speciation.* (pp. 57–75). New York: Oxford University Press.

de Queiroz, K. (2007). Species concepts and species delimitation, *Systematic Biology, 56,* 879–886. doi:10.1080/10635150701701083

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. doi:10.1371/journal.pone.0104982

Ducasse, J., Ung, V., Lecointre, G., Miralles, A. (2020). LIMES : a tool for comparing species partition. *Bioinformatics*, 2282–2283. doi: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. doi: 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. doi:10.1093/sysbio/syy072.

Ence, D.D., & Carstens, B.C. (2011). SpedeSTEM: A rapid and accurate method for species delimitation. *Molecular Ecolology Resources, 11*, 473–480. doi: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. Biological Sciences.* 279(1743), 3678–3686. doi:10.1098/rspb.2012.0705.

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. doi:10.1186/1471-2148-10-372

Flot, J.-F. (2015). Species delimitation's coming of age, *Systematic Biology, 64,* 897–899.

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. doi:10.1371/journal.pbio.0050087

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. doi : 10.1093/sysbio/syt033

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. doi:10.1093/sysbio/syw028

García-Melo, J. E., Oliveira, C., Da Costa Silva, G. J., Ochoa-Orrego, 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. doi: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. doi: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. Ney York, Basel.

Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology, 74,* 447. doi: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. doi: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. doi:10.1093/bioinformatics/btx025

Knowles, L. L., Carstens, B. C. (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, 56 (6), 887–895. doi :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 *Recent Advances in Lichenology* (pp. 11–44). New Delhi: Springer. doi: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. doi: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. doi: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. doi: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. doi:10.1111/j.1755-0998.2010.02896.x

Miralles, A. & Vences, M. (2013). New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PlosONE, 8,* e68242. doi:10.1371/journal.pone.0068242

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., Wilke, T., Renner, S., Vences, M. (2020). Repositories for taxonomic data: where we are and what is missing. *Systematic Biology, 69,* 1231–1253. doi:10.1093/sysbio/syaa026

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. doi: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. doi: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. doi: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. doi:10.1016/j.ympev.2016.08.013

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012b). ABGD, Automatic Barcode Gap Discovery for primary species delimitation, *Molecular Ecology, 21,* 1864–1877. doi: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. (2012a). Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology, 11,* 2671–3691. doi:10.1111/j.1365-294X.2012.05559.x

Puillandre, N., Brouillet, S., Achaz, G. (2021). ASAP: Assemble Species by Automatic Partitioning. *Molecular Ecology Resources*, 21(2), 609–620. doi:10.1111/1755-0998.13281

Rabiee, M., Mirarab, S. (2019). SODA: Multi-locus species delimitation using quartet frequencies. *bioRxiv,* 869396. doi:10.1101/869396

Rannala, B. (2015). The art and science of species delimitation. *Current Zoology*, 61, 846–853. doi: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. doi: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. doi: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. doi:10.1111/j.1095-8312.2006.00689.x

Sites, J. W., & Marshall J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution, 18,* 462–470. doi:10.1016/S0169-5347(03)00184-8

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., Keith, P., Rüber, L., Hubert, N., (2020). Disentangling the taxonomy of the subfamily Rasborinae (Cypriniformes, Danionidae) in Sundaland through DNA barcodes**.** *Scientific Reports,* 10, 2818. doi:10.1038/s41598-020-59544-9

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. doi: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. doi:10.1111/2041-210X.13454

Sukumaran, J., & Knowles, L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences USA*, 114(7), 1607–1612. doi:10.1073/pnas.1607921114

Sukumaran, J., Holder, T. M., Knowles, L. L. (2021). Incorporating the speciation process into species delimitation. *PloS Computional Biology* 17(5): e1008924. doi :10.1371/journal.pcbi.1008924

Vences, M. (2020). The promise of next-generation taxonomy. *Megataxa, 1,* 35–38. doi:10.11646/megataxa.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. (in press). iTaxoTools 0.1: Kickstarting a specimen-based software toolkit for taxonomists. *Megataxa*.

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 U. S. A., 106,* 8267–8272. doi: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. doi: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. doi:10.1080/10635150275347588O

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences U. S. A., 107,* 9264–9269. doi: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. doi:10.1093/molbev/msu279

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics, 29,* 2869–2876. doi:10.1093/bioinformatics/btt499

Zhang, C., Zhang, D. X., Zhu, T., Yang, Z. (2011). Evaluation of Bayesian coalescent method of species delimitation. *Systematic Biology,* 60(6), 747–761. doi:10.1093/sysbio/syr071

Data Accessibility 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 iTaxotools website (http://itaxotools.org).

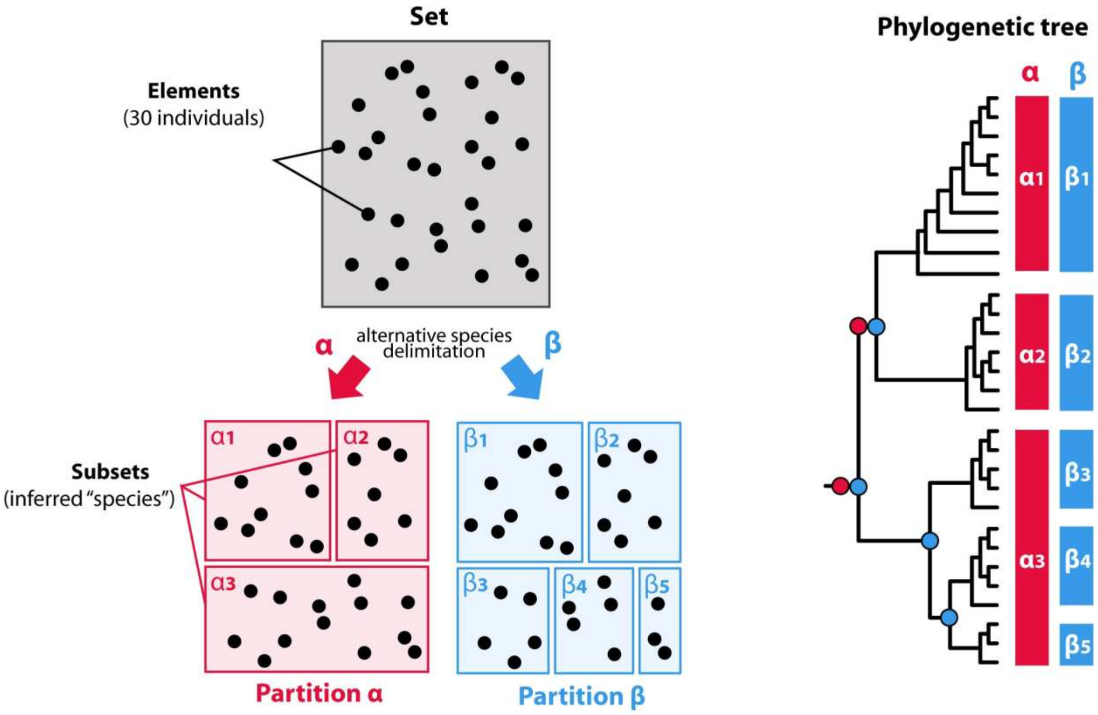


Figure 1. In mathematics, a partition of a set is a grouping of its elements into non-empty subsets, in such a way that every element is included in exactly one such subset. The main output of a species delimitation inference therefore corresponds to a partition, independently of the theoretical context, 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 α and β, corresponding to 3 and 5 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 color. Note that not all SD methods rely on a tree topology, and may therefore delimit non-monophyletic units (e.g., methods based on morphological or molecular divergence).

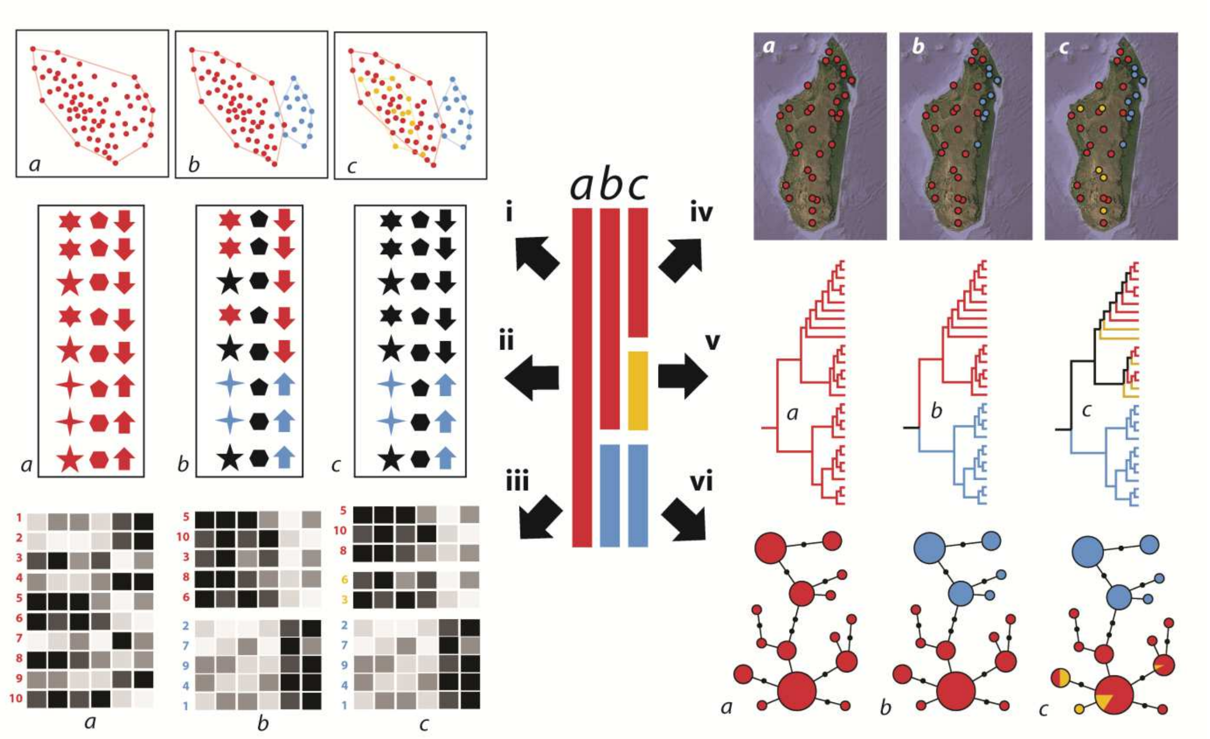
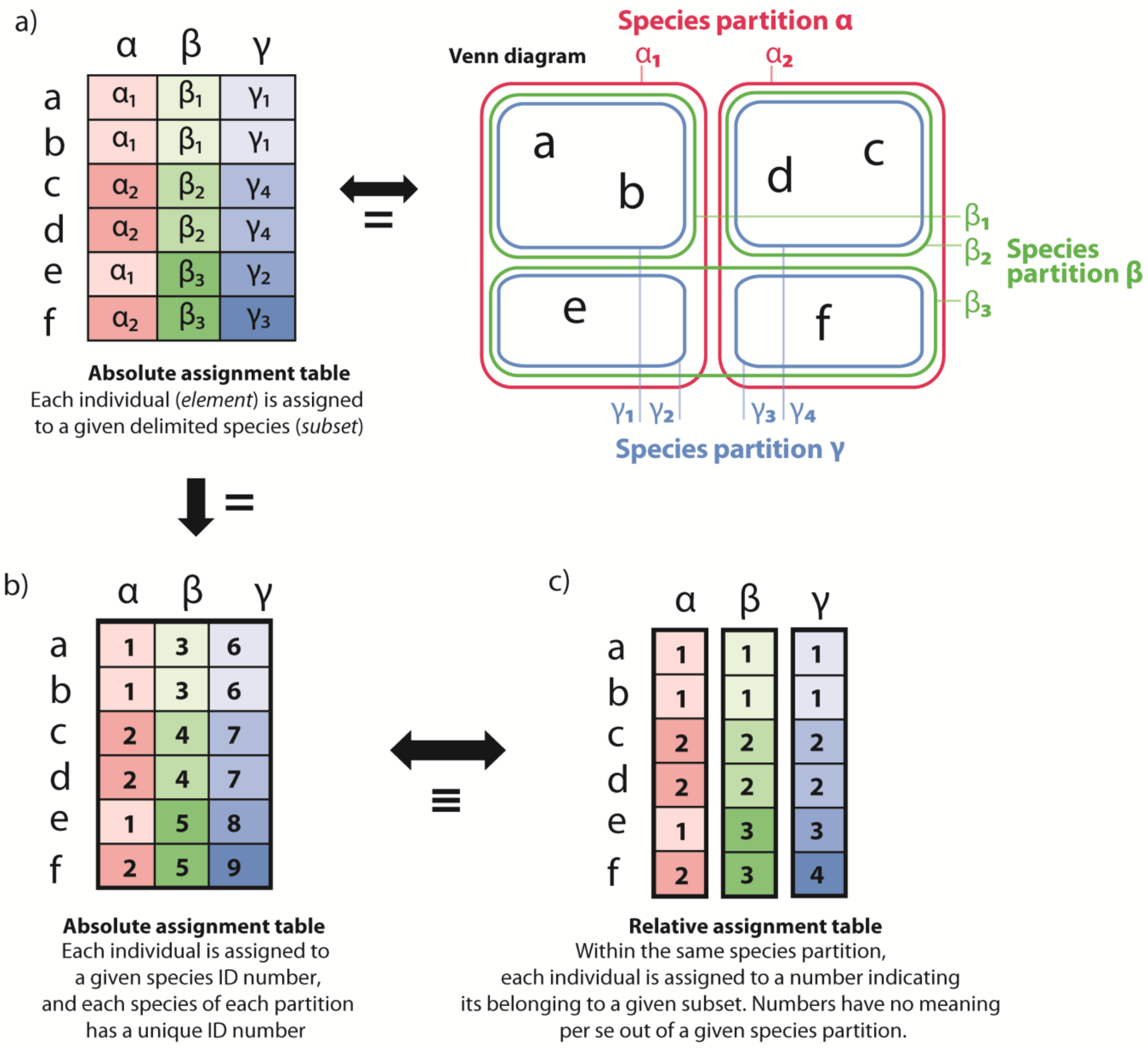


Figure 2. Illustration of exemplary potential applications of a species partition (SPART) file. If it can be parsed by other programs, SPART might facilitate the exploration of taxonomic datasets under various delimitation assumptions (such as (i) morphometric Principal Component Analysis, (ii) automated extraction of diagnostic traits (three qualitative morphological characters with various states in this example), (iii) heatmap of meristic morphological traits (for a visual exploration of the phenotypic variability), (iv) distribution map, (v) mitochondrial DNA-based phylogenetic tree, or (vi) haplotype network from nuclear DNA). 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.



	α	β	γ
a	α_1	β_1	γ_1
b	α_1	β_1	γ_1
c	α_2	β_2	γ_4
d	α_2	β_2	γ_4
e	α_1	β_3	γ_2
f	α_2	β_3	γ_3

Absolute assignment table
Each individual (*element*) is assigned
to a given delimited species (*subset*)

	α	β	γ
a	1	3	6
b	1	3	6
c	2	4	7
d	2	4	7
e	1	5	8
f	2	5	9

Absolute assignment table
Each individual is assigned to
a given species ID number,
and each species of each partition
has a unique ID number

	α	β	γ
a	1	1	1
b	1	1	1
c	2	2	2
d	2	2	2
e	1	3	3
f	2	3	4

Relative assignment table
Within the same species partition,
each individual is assigned to a number indicating
its belonging to a given subset. Numbers have no meaning
per se out of a given species partition.

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 (α, β and γ). All these species partitions are hierarchically compatible (i.e. they conform to the mathematical definition of nested sets), with the exception of the pair α - β (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).

Table 1. Automated tools dedicated to species delimitation. Abbreviations used: mtDNA, mitochondrial DNA; nDNA, nuclear DNA. Note that for programs marked with an asterisk (GMYC, PTP, 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 assignement)	Optimal datasets and format	SPART impletementation	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 & Rannala (2010, 2014), Solís-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 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 & Hebert (2013)
PTP (mPTP and bPTP)	Multi-rate Poisson ree processes model	No	Non ultrametric 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 & Mirarab (2019)
HaplowebMaker / CoMa	Mutual allelic exclusivity	No	nDNA – multilocus alignments	No	Spöri & 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. (2020)

SPART, a versatile and standardized data exchange format for species partition information.

Aurélien Miralles, Jacques Ducasse, Sophie Brouillet, Tomas Flouri, Tomochika Fujisawa, Paschalia Kapli, L. Lacey Knowles, Sangeeta Kumari, Alexandros Stamatakis, Jeet Sukumaran, Sarah Lutteropp, Miguel Vences, Nicolas Puillandre

Appendix 1: Technical description of the Spart format (species partition)

(version 12/03/2021)

1. Background, environment and software implementation page 2

2. Terminology page 4

3. Technical details page 5

4. Format description: matricial SPART (Nexus-like) page 5

5. Format description: SPART.XML page 12

Participants in the initial development of the spart format:
Brouillet S. (programming for ABGD, ASAP and other tools of the future MNHN platform),
Ducasse J. (programming for LIMES),
Kumari S. (programming for iTaxoTool which will implement various SD methods)
Miralles, A (design, MNHN),
Puillandre N. (design, MNHN),
Vences, M. (design, TU Braunschweig)

1. Background, environment and software implementation

The spart formats

Two implementations of the spart format are proposed:

- ► a matricial spart format (SPART) in which for each individual (sample), multiple species partition assignment is included in a concatenated, table-like format. This format has been designed to be intuitively understandable by humans, facilitating manual editing and import into table editors, and has a syntax largely compatible with the nexus format, commonly used in phylogenetics, thus facilitating its inclusion as a separate block into nexus files if required by future analysis software.
- ► a SPART.XML format in which the information for each species partition is provided in a separate block, and in which blocks are serially appended one after each other. This format is optimized for being machine-readable and its syntax follows the XML language.

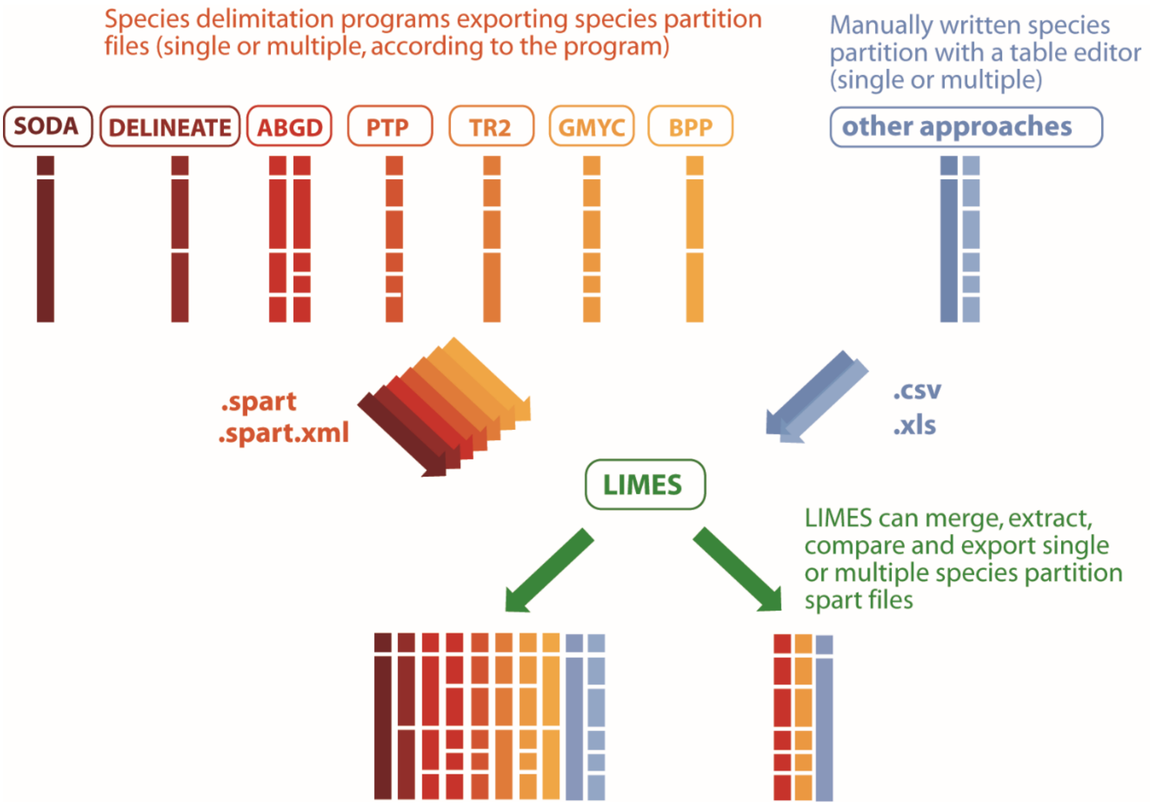
We strongly recommend using the extensions ".spart" and "spart.xml" for the matricial and XML implementation, respectively.

Current implementation (march 2021):

Programs currently implementing SPART : ABGD, ASAP, GMYC, PTP, SODA, TR2, DELINEATE, LIMES.

LIMES v2.0 (http://itaxotools.org/download.html) will act as a central platform for converting and modifying spart files.
LIMES v2.0 is compatible with matricial spart (.spart) files; spart.xml compatibility will be implemented in the next version, together with the possibility to convert between matricial SPART and SPART.XML files. In addition, standalone and web-based tools will be implemented in the future to easily convert spart from and to spart.xml files.

LIMES v2.0 can read one or several species partitions from a CSV formatted document (thus including manually created spartitions), merge species partitions (= spartitions) from several single and/or multiple spartitions files, extract them, and export them into a single multi-spartitions spart file.


Species delimitation programs exporting species partition files (single or multiple, according to the program)
Manually written species partition with a table editor (single or multiple)
SODA
DELINEATE
ABGD
PTP
TR2
GMYC
BPP
other approaches
.spart
.spart.xml
.csv
.xls
LIMES
LIMES can merge, extract, compare and export single or multiple species partition spart files

2. Terminology

Species partition (Spartition): Distribution (classification or assignment) of all individuals into multiple subsets, according to a given method.

Subset: elementary unit of the results (of the partition); usually a “species”, but can also be defined as a cluster or an operational taxonomic unit (OTU) or a molecular operational taxonomic unit (MOTU) or a barcode index number (BIN), a population (e.g. STRUCTURE) or any other kind of unit, depending on the computational analysis performed.

Individual: elementary unit of the dataset; usually equals a sample or a specimen in the SD analysis which in most cases will represent an individual organism (but can also be for instance an isolate/culture in microbiology).

Spartition score: any score attributed to the species partition as a whole (one score per species partition, usually corresponding to one score per SD analysis).

Subset score: any score attributed to each subset (to quantify its distinctiveness relative to the others).

Individual score: any score attributed to the assignment to a subset proposed for each individual.

Note: inclusion of these three scores in a spart file is optional, but if any such scores are calculated by an SD software, we recommend that the output spart files should include this information in the proposed format.

Single species partition file: Usually, the result of a species delimitation analysis (SD). Most often, each SD program is expected to export a file with a single species partition, but exceptions exists (e.g., ABGD typically provides several partitions and thus may either export multiple single species partition files, or one multiple species partition file).

Multiple species partition file: the information of several single species partitions merged into a single file, either as direct output from some SD programs, or by merging single species partition files using LIMES (or other tools with this functionality). The respective developers of most SD programs will typically implement the export of a single species partition file per analysis.

Note: ABGD and ASAP are already producing multiple species partitions as a result of a single analysis. So ideally, the user should be allowed to select which species partitions to export at the end of the analysis, and whether this should be done as multiple single species partition files, or as a single multiple species partition file. Typically, both ABGD and ASAP will provide a list of partitions, but some of them are often unrealistic (especially for ABGD, when they are far from the barcode gap), and the user may not desire to include them in the spart file exported by the program and used for further analysis.

Block: To enable compatibility with programs using Nexus as input file, the matricial spart file is conceived as a single block. Its start is indicated by an initial line specifying "begin spart;" and its end is indicated by a line specifying "end;".
In the spart.XML format, all information of one species partition is provided as one block separate from other blocks (species partitions).

Command (matricial SPART format): A command corresponds to a section/field intended to provide a specific type of information or instruction (e. g. “N_spartitions”, “N_subsets”, “N_individuals”, “Individual_assignment”, including the subsequently given details and values, each correspond to a different command).

Command title: Specific title given to a given command (e. g. “N_spartitions”, “N_subsets”, “N_individuals”, “Individual_assignment”).

4. Format description: matricial spart

1) Character set :

By default: All the 95 ASCII printable characters are allowed in the entire format (incl. space)

``!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMN``

``OPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}``

Exception for individual (sample) names (assignment list): only numbers, capital and lower case letter and underscore (no spaces nor any other diacritic signs) are allowed:
`0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ_abcdefghijklmnopqrstuvwxyz`
We recommend to use underscore to replace any forbidden character

Any individual name including another character should lead to a specific error message

Exception for partition names (ex. in N_spartitions):
These six characters are strictly prohibited :
~~comma~~ (,)
~~colon~~ (:)
~~slash~~ (/)
~~semi colon~~ (;)
~~opening and closing brackets~~ ([ ) ( ])

Any partition name including one of these six characters should lead to a specific error message. More generally, we recommend to only use numbers, capital and lower case letters (i.e. to use underscores to replace any other symbols), as above (individual samples).

Exception for species assignment (Individual_assignment) : only positive integers are allowed (`0,1, 2, 55, 101, 102`, etc.)

Any assignment using another character should lead to a specific error message

2) Scores

Scores will often represent a proportion (between 0 and 1) such as posterior probabilities or bootstrap proportions, but can also take very small values (e.g., likelihood scores).

Spartition, species and individual **scores** are either in the form of fixed-point notation (e.g. 0.093) or in the form of floating point numbers (scientific (exponential) notation (e.g., 9.30E-02)).

Negative values are permitted (e.g. log likelihood values)

Question marks (?) represent missing data.

A score category can be removed if (and only if) totally empty: in particular (but not limited to) cases where there is no spartition score at all, or no subset score at all, or no individual score at all.

3) Separators

Spart should be robust against whitespaces (we prefer to not take any risk, although Nexus separates e.g. taxon names and characters by whitespaces). The spart format avoids using tabs or spaces as separators. **The only separators allowed are the following:**

Colons (:) first order separator (although the term separator is not fully appropriate here). Colons are used to separate an item followed by a list of attributes sharing the same order.

Slashes (/) second order separator, used to separate values corresponding to different species partitions.

Comma (,) third order separator, used only for partition scores and subset scores.

Equality signs (=), separating a Command title from the values presented after the sign.

Additional “pseudo-separators”:

End of line: to end every line (within a given command) referring to an individual sample. Each new command (section) starts by a new line. End of line are NOT allowed within an individual sample line (ex. with a line of “Individual_assignement).

It is important that programs reading matricial spart files accept all possible end of lines, i.e., Windows (CR-LF), Unix (LF) and old Mac (CR)

Semi-colon: to indicate the end of a command. Only the semi-colon indicates “end of the command”. That means that it can appear after the last line, or in the line afterwards. The two following examples are equivalent and correct:

```
N_spartitions = 3:
     CO1_ABGD, 0.98 /
     test_BPP, 0.95 /
     PCA_phenotype,?
;
```

```
N_spartitions = 3:
     CO1_ABGD, 0.98 /
     test_BPP, 0.95 /
     PCA_phenotype,?;
```

Brackets [in order to isolate a comment like in this sentence].
A comment can appear embedded within a command (between a command title (beginning) and the end (;)). As brackets are used to frame a comment, a comment cannot contain a bracket embedded within it. Exemple: [this comment is ~~[not]~~ correct]

4) Syntax

Spaces: should have no influence (Spart should be robust against whitespaces).
For example : N_individuals = 5 / 5, N_individuals =5/5 , and N_individuals =5/   5 should all be correct and equivalent (i.e software should be able to read a file written by hand and containing minor errors like these above).

Nevertheless, we recommend to implement (to automatically generate e.g. as output of SD programs) only the following format: **N_individuals = 5 / 5**

Commands (N_spartitions, N_subsets, N_individuals, Individual_assignment…): are not case sensitive and the following examples all are equivalent and should be readable by programs: N_SUBSETS, n_subsets or N_SUBsets

We recommend to implement (to automatically generate) only the following format: **N_subsets**

Command order:
The order of the compulsory commands must be respected:
1: Project_name, 2: Date, 3: N_spartitions, 4: N_individuals, 5: N_subsets, 6: Individual_assignment
The optional commands must appear after the compulsory commands, but their respective order is free.

Begin and end of spart file (spart block):

Because in the matricial spart format, all information resides in one block, the beginning of the block is specified at the very beginning of the file (i.e., before the Project_name command):

```
begin spart;
```

and the end is specified either at the very end of the file (either after the last compulsory command if only compulsory commands are included in the file; or after the last optional command):

```
end;
```

Commands in the matricial spart format (SPART), exemplified by a multiple partition file
(but note that most SD programs will usually export a single species partition file)

Compulsory commands	**These commands need to be present in any spart file.** If any of them is missing, some programs using the spart file will possibly not work or output error messages.-> ERROR MESSAGE
begin spart;	Starting line of the spart file. Indicates the begin of a block (the entire spart file is conceived as a single block).
Project_name = my_three_delimitations;	Name given for this new project

Date = 2020-09-21T07:26:10+00:00;	Date and time (standard ISO 8601 recommended) in which **this** specific spart file was generated. The date is mandatory but the format is flexible.
	These three examples are correct :
	Date=2021-03-04T16:35:30.767494+01:00 ;
	Date=2020-09-21T07:26:10+00:00 ;
	Date=2020-09-21T07:26:10 ;
	Date=2020-09-21 ;

N_spartitions = 3:CO1_ABGD, 0.98 / test_BPP, 0.95 / PCA_phenotype, ? ; The spartition scores can also be omitted (e.g. if no score at all), in which case the command would read: **N_spartitions =** 3:CO1_ABGD / test_BPP / PCA_phenotype;	Number of species partitions; list of spartitions names , **spartition score (? If no score)** Spartition names are separated by slashes. This command define the order the spartitions (1st=CO1, 2nd=BPP, 3rd=PCA) that will be reused in the subsequent commands. Two different spartitions are not allowed to share the same name [note: spartition scores are included with number and names of spartitions and not in an optional additional command to facilitate extraction of information by human readers]
N_individuals = 5 / 5 / 4;	Total number of individuals (=samples, specimens) (1st, 2nd and 3rd spartition) (the spartition order is the same as in N_spartitions)
N_subsets = 3:0.95,0.98,0.99 / 2:0.95,0.98 / 4:?,?,?,?; The subset scores can also be omitted, in which case the command would read: **N_subsets =** 3 / 2 / 4;	Total number of delimited subsets (1st, 2nd and 3rd spartition), subset score (? If no scores) [note 1: The first score corresponds to the first subset appearing in the Individual assignment list (from top to bottom), the second subset score correspond to the second subset appearing in the list, etc. *Therefore, it is a "top to bottom" order based on individual assignment list, independently* from the *"value" of the number used to assign an individual to a given subset.* [note 2: subset scores are included after the number of subsets]
[CO1_ABGD : this is my first comment**]**	[Comment]

[CO1_ABGD : this is my second comment] [PCA_phenotype : this is my first comment extracted from the third method] [my_three_delimitations : possible comment related to the concatenated multiple partition file]	A comment begins with an opening bracket and finish with a closing bracket. A multiple (concatenated) species partition file is able to report all the comments made independently in the different single species partition file (SPF), so the name of each SPF should be reported at the beginning of each comment. [note: A comment, if needed, can be place anywhere in the file (see below). It can be in a single line (ex. both comment of CO1_ABGD) or on different lines (ex. PCA_Phenotype)]
Individual_assignment = Drosophila_32:1/1/4 [a comment can be placed anywhere] Sample_2:1/1/3 Drosophila_China:2/2/2 Sample_E554:2/1/? Droso_Vietnam:3/2/1; [CO1-ABGD : comment about the CO1 assignment] [PCA-ABGD : comment about the PCA assignment]	List of individuals (samples) with their respective assignment in each of the three spartitions (1st, 2nd , then 3th spartitions), i.e., usually by each of three methods (? If individual not assigned by one of these methods). Two different samples are not allowed to share the same name. End of lines are separating each individual line : Drosophila_32:1/1/4 ⏎ Sample_2:1/1/3⏎ Drosophila_China:2/2/2⏎ Sample_E554:2/1/?⏎ Droso_Vietnam:3/2/1;
end;	Ending line of the spart file. Indicates the end of a block. (the entire spart file is conceived as a single block). This line must be at the every end of the spart file (i.e., after the very last included command (wether they are compulsory or optional).
Optional commands	These optional fields are not part of the basic, compulsory spart syntax. They may be present in the spart files (and will be carried over or specifically generated if various spart files are merged into one), but **if they are missing it should not generate an error message**, except in such programs that specifically expect/require the information of some of these optional fields. Remember: In general terms, the spart readers/parsers/analyzers should work in a way that they simply ignore lines with information they do not "understand" so that it becomes easy to add additional optional fields if it is later deemed to be useful for some specific applications.
Individual_score = Drosophila_32:?/0.99/1.00 Sample_2:?/?/1.00 Drosophila_China:?/0.97/0.99 Sample_E554:?/0.85/? Droso_Vietnam:?/0.99/0.96;	List of individuals (samples) with their respective individual **score** according to each method, ie. 1st, 2nd then 3rd spartitions (? if no score). Optional command: no need to present this command, e.g. if it is totally without values. End of lines are separating each individual line : **Individual_score =** Drosophila_32:?/0.99/1.00 ⏎ Sample_2:?/?/1.00⏎

	Drosophila_China: ? / 0.97 / 0.99⏎ Sample_E554: ? / 0.85 / ? ⏎ Droso_Vietnam: ? / 0.99 / 0.96; It is important to accept either Windows (CR-LF), Unix (LF) and old Mac (CR) end of line [note: individual scores are given as separate (optional) command and are not included in the Assignment command because often they will be missing altogether, and because presenting them separately facilitates extraction of information in the Assignment command by human readers]
Spartition_score_type = likelihood / ? / ?; **Subset_score_type** = bootstrap / ? / posterior_probability; **Individual_score_type** = probability / bootstrap / ?;	If these commands are absent, then the respective score types are missing = "?" [note: Score types are flexibles (no defined list of type)]
Tree = test_BPP : ((Drosophila_32, Sample_2), Drosophila_China,(Sample_E554, Droso_Vietnam)) CO1_ABGD : ((Drosophila_China, Sample_2), Drosophila_32,(Sample_E554, Droso_Vietnam)) ;	Reports (chain of characters) the input tree used for the calculation of a certain partition, in Newick format. Multiple trees can be included, especially in a multi-spart file (one or maybe even several for each partition)
Command_line = test_BPP : 2019-01-30T09:26:10+00:00 / BPP version 2.0 / "bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla” CO1_ABGD : 2019-01-30T09:26:10+00:00 / 3.0 available at abgd.com / "abgd myinputfile.fas -a -v -P 0.3" ;	Gives the full command line of the program that was executed for generating the delimitation for the respective species partition (with the date, if existing)). **Commandline space** = name of the respective species partition : date of the original analysis / version of the tool used / specific commandline that was executed (in quotation marks) Question marks in case of (partly) missing data. **semicolon** (to end the line).

5. Format description: SPART.XML

This format follows the Extensible Markup Language (XML), a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

The spart.xml format encodes the same information as the matricial spart, but with a vocabulary adapted to fit conventions and requirements of XML. In particular this affects the following commands:
spartition_score = spartitionScore
individual_score = individualScore
individual_score_type = individualScoreType
subset_score = subsetScore

spartition_score_type and subset_score_type are not used as separate terms, but the respective information encoded in the respective lines subsetScore and and spartitionScore under "type"; the respective values are given in the same lines under "value". Subset scores are furthermore placed in a section "external support" which can also provide information ("source") on the type of analysis, algorithm or program this support was derived from.

```
<?xml version="1.0" ?>
<root>
          <project_name>Mantella.fas</project_name>
          <date>2021-01-29T18:13:35</date>
    <!-- Generated by bPTP -->
    <!-- WARNING: The sample names below may have been changed to fit SPART specification (only alphanumeric characters and _ ) -->
    <!-- user comment: this analysis was generated based on a single ML tree obtained in MEGA 7 -->
          <individuals>                                                                                                List of individuals sampled
        <individual id="aura_ZCMV1234"  />
        <individual id="aura_ZCMV1235"  />
        <individual id="aura_ZCMV1236"  />
        <individual id="aura_ZCMV1237"  />
        <individual id="aura_ZCMV1238"  />
        <individual id="aura_ZCMV1239"  />
        <individual id="aura_FGZC987"   />
        <individual id="aura_FGZC986"   />
        <individual id="crocea_ZCMV234" />
        <individual id="crocea_ZCMV235" />
        <individual id="miloty_ACZC324" />
        <individual id="miloty_ACZC329" />
        <individual id="crocea_ZCMV236" />
        <individual id="crocea_ZCMV237" />
        <individual id="miloty_ACZV679" />
        <individual id="miloty_ZCMV479" />
          </individuals>
          <spartitions>                                                                                             Spartitions description

                   <spartition label="Mantella_bPTP" spartitionScore="1.234E-6" spartitionScoreType="logLikelihood" > First spartition description

            <remarks>First spartition</remarks>
                             <subsets>                                                                 Subsets description of the first spartition (N=3 in this exemple)
                                      <subset label="1">
                                          <externalSupport>
                                              <subsetScore type="posterior" value="1.23E-6" source="BEAST analysis 2021-03-02" />
                                          </externalSupport>
                                                <individual ref="aura_ZCMV1234" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_ZCMV1235" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_ZCMV1236" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_ZCMV1237" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_ZCMV1238" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_ZCMV1239" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_FGZC987"  individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="aura_FGZC986"  individualScore="1.23E-3" individualScoreType="probability" />
                                      </subset>
                                      <subset label="2">
                                          <externalSupport>
                                              <subsetScore type="posterior" value="7.34E-6" source="BEAST analysis 2021-03-02" />
                                           </externalSupport>
                                                <individual ref="crocea_ZCMV234" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="crocea_ZCMV235" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="miloty_ACZC324" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="miloty_ACZC329" individualScore="1.23E-3" individualScoreType="probability" />
                                                <individual ref="crocea_ZCMV236" individualScore="1.23E-3" individualScoreType="probability" />
```

```
                                                 <individual ref="crocea_ZCMV237" individualScore="1.23E-3" individualScoreType="probability" />
                                                 <individual ref="miloty_ACZV679" individualScore="1.23E-3" individualScoreType="probability" />
                                                 <individual ref="miloty_ZCMV479" individualScore="1.23E-3" individualScoreType="probability" />
                                       </subset>
                                       <subset label="3">
                                           <externalSupport>
                                               <subsetScore type="posterior" value="1.01E-5" source="BEAST analysis 2021-03-02" />
                                           </externalSupport>
                                                 <individual ref="miloty_ACZV679" individualScore="1.23E-3" individualScoreType="probability" />
                                                 <individual ref="miloty_ZCMV479" individualScore="1.23E-3" individualScoreType="probability" />
                                       </subset>
                             </subsets>
                    </spartition>
                    <spartition label="analysis P2"> <score type="likelihood" value="1.0345E-06" />
            <remarks>Second spartition</remarks>

                             <subsets>
                                       <subset label="1">
                                                 <individual ref="aura_ZCMV1234" />
                                                 <individual ref="aura_ZCMV1235" />
                                                 <individual ref="aura_ZCMV1236" />
                                                 <individual ref="aura_ZCMV1237" />
                                       </subset>
                                       <subset label="2">
                                                 <individual ref="aura_ZCMV1238" />
                                                 <individual ref="aura_ZCMV1239" />
                                       </subset>
                                       <subset label="3">
                                                 <individual ref="aura_FGZC987" />
                                                 <individual ref="aura_FGZC986" />
                                       </subset>
                                       <subset label="4">
                                                 <individual ref="crocea_ZCMV234" />
                                                 <individual ref="crocea_ZCMV235" />
                                                 <individual ref="miloty_ACZC324" />
                                                 <individual ref="miloty_ACZC329" />
                                                 <individual ref="crocea_ZCMV236" />
                                                 <individual ref="crocea_ZCMV237" />
                                                 <individual ref="miloty_ACZV679" />
                                                 <individual ref="miloty_ZCMV479" />
                                       </subset>
                                       <subset label="5">
                                                 <individual ref="miloty_ACZV679" />
                                                 <individual ref="miloty_ZCMV479" />
                                       </subset>
                             </subsets>
                    </spartition>
          </spartitions>
</root>
```

Second spartition description

Subsets description of the 2nd spartition (N=5 in this exemple)