



**HAL**  
open science

## TreePics: visualizing trees with pictures

Nicolas Puillandre, Julien Brisset, Simon Chagnoux, Etienne Cavillier

► **To cite this version:**

Nicolas Puillandre, Julien Brisset, Simon Chagnoux, Etienne Cavillier. TreePics: visualizing trees with pictures. *European Journal of Taxonomy*, 2017, 353, pp.1-7. 10.5852/ejt.2017.353 . hal-01617897

**HAL Id: hal-01617897**

**<https://hal.sorbonne-universite.fr/hal-01617897>**

Submitted on 17 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This work is licensed under a Creative Commons Attribution 3.0 License.

## Opinion paper

[urn:lsid:zoobank.org:pub:6EC2A7CA-6704-4CDF-A204-F02A39B6AE18](https://doi.org/10.5852/ejt.2017.353)

# TreePics: visualizing trees with pictures

Nicolas PUILLANDRE<sup>1,\*</sup>, Julien BRISSET<sup>2</sup>, Simon CHAGNOUX<sup>3</sup> & Etienne CAVILLIER<sup>4</sup>

<sup>1</sup>Institut de Systématique, Evolution, Biodiversité ISYEB – UMR7205 – CNRS, MNHN, UPMC, EPHE, Muséum national d’Histoire naturelle, Sorbonne Universités, 43 rue Cuvier, F-75231 Paris, France.

<sup>2</sup>Muséum national d’Histoire naturelle, UMS 2006 Patrimoine Naturel, 36 rue Geoffroy Saint-Hilaire, CP41 75005 Paris, France.

<sup>3,4</sup>Muséum national d’Histoire naturelle, Division des Systèmes d’Information, 57 rue Cuvier, CP41 75005 Paris, France.

\* Corresponding author: [puillandre@mnhn.fr](mailto:puillandre@mnhn.fr)

<sup>2</sup>Email: [brisset@mnhn.fr](mailto:brisset@mnhn.fr)

<sup>3</sup>Email: [simon.chagnoux@mnhn.fr](mailto:simon.chagnoux@mnhn.fr)

<sup>4</sup>Email: [etienne.cavillier@mnhn.fr](mailto:etienne.cavillier@mnhn.fr)

<sup>1</sup>[urn:lsid:zoobank.org:author:00565F2A-C170-48A1-AAD9-16559C536E4F](https://doi.org/10.5852/ejt.2017.353)

<sup>2</sup>[urn:lsid:zoobank.org:author:D9B88176-125B-4D0E-BAEF-2A99A879BD3A](https://doi.org/10.5852/ejt.2017.353)

<sup>3</sup>[urn:lsid:zoobank.org:author:802DE9F8-B428-4A1E-A068-1A2B043D3BA2](https://doi.org/10.5852/ejt.2017.353)

<sup>4</sup>[urn:lsid:zoobank.org:author:E7E22101-C771-4C66-84C5-5E2E5480957C](https://doi.org/10.5852/ejt.2017.353)

**Abstract.** While many programs are available to edit phylogenetic trees, associating pictures with branch tips in an efficient and automatic way is not an available option. Here, we present TreePics, a standalone software that uses a web browser to visualize phylogenetic trees in Newick format and that associates pictures (typically, pictures of the voucher specimens) to the tip of each branch. Pictures are visualized as thumbnails and can be enlarged by a mouse rollover. Further, several pictures can be selected and displayed in a separate window for visual comparison. TreePics works either online or in a full standalone version, where it can display trees with several thousands of pictures (depending on the memory available). We argue that TreePics can be particularly useful in a preliminary stage of research, such as to quickly detect conflicts between a DNA-based phylogenetic tree and morphological variation, that may be due to contamination that needs to be removed prior to final analyses, or the presence of species complexes.

**Keywords.** Phylogenetic trees, voucher pictures, DNA barcoding, computer program.

Puillandre N., Brisset J., Chagnoux S. & Cavillier E. 2017. TreePics: visualizing trees with pictures. *European Journal of Taxonomy* 353: 1–7. <https://doi.org/10.5852/ejt.2017.353>

## Introduction

The DNA-barcoding initiative (Hebert *et al.* 2003) and Next-Generation Sequencing technologies have dramatically increased the amount of molecular data available for biodiversity studies. It is now common

to manipulate datasets of 100s or even 1000s of haplotypes for a given genetic marker. Typically, DNA barcodes obtained from several specimens in several species are compared to test whether intra- and interspecies diversity matches (or not) the accepted species hypotheses, which are generally based on morphological characters. Such comparisons will eventually lead to detecting highly structured groups within a given species, then used as primary species hypotheses in an integrative taxonomy approach (Dayrat 2005).

There is a need for efficient bioinformatic tools to handle increasingly larger datasets, and many tools have been proposed to turn raw data in a wide range of input formats for analytic software. One critical step is to detect potential conflicts between DNA sequences and the phenotypes of the specimens from which the sequences were obtained (the vouchers). Such conflicts could actually correspond to “true” discrepancies, i.e., several morphs correspond to a single haplotype (or closely related and highly similar haplotypes) due to, e.g., rapid adaptive radiation, incomplete lineage sorting or phenotypic plasticity, or conversely, one morph actually encompasses several molecular units due to, e.g., convergence or retention of ancestral polymorphism (e.g., Sangster 2009). However, contamination by other living organisms can also happen, even if evidence for that in the literature is not common (see, e.g., Schmidt *et al.* 1995): negative results are rarely published and in most cases contamination is removed prior to final analyses. While some contaminations are easy to detect (e.g., contamination by very distant organisms), others are more tricky to identify (e.g., contamination by a closely related species). The BLAST tool in GenBank is commonly used to detect such contaminations, but many species are not represented in GenBank, and detecting contaminations often requires comparing sequences from closely related species. Thus, one way to efficiently detect such discrepancies is to visualize the data with a phylogenetic tree and compare the phenotypes of the sequenced specimens in the order they appear in the tree. While numerous standalone programs are available to visualize phylogenetic trees (e.g., FigTree (Rambaut & Drummond 2009); Treeview (Page 2002); MrEnt (Zucon & Zucon 2010);...) none of them offer, to our knowledge, the possibility to automatically display pictures associated to the sequenced specimens in a way that enables a direct comparison of a tree and hundreds of associated images. In our case, this gap in the analytical workflow appeared when we started to accumulate barcoding data from the specimens preserved in the collection of the MNHN (Muséum national d’Histoire naturelle) and stored in its databases (Puillandre *et al.* 2012). Sample contamination occurred regularly because we were working with universal primers on specimens that were potentially not perfectly preserved and collected over several decades. To allow a rapid identification of conflicts between DNA phylogenies and morphological characters, we present the program TreePics as a webtool and as a standalone software. TreePics can be used to visualize a tree in Newick format and, at the tip of each branch, a picture corresponding to the specimen (typically, a picture of the voucher, but potentially any picture related to it, a detail of the voucher, an icon related to a phenotypic trait,...). It is not designed to edit a tree, and can thus be used in combination with other programs that are dedicated to tree editing. The ETE toolkit (Huerta-Cepas *et al.* 2016) is also able to associate pictures to branch tips, but it needs to be used within a Python framework, where pictures and taxa names are associated manually. TreePics has been designed to automatically associate hundreds or thousands of pictures to a phylogenetic tree, using only a web browser.

### **General functioning**

TreePics is accessible through a web browser. It is based on the web standards HTML5, CSS3, javascript and the library jQuery. The input tree in the Newick format is converted in the SVG (Scalable Vector Graphics) format using the library jsPhyloSVG (Smits & Ouverney 2010) to obtain a vector graphic representation of the tree which is understood natively by modern browsers.

All operations in the web browser are performed in javascript, and all the operations are done locally in the web browser. TreePics is thus independent from any application server, and the pictures do not need

to be uploaded (either on the server or on cloud storage). Because the size of the pictures can easily reach several gigabytes, this step would have been highly time- and space-consuming.

If a web application is not constrained by the heterogeneity of the operating systems, it must nonetheless take into account the security rules imposed by the web browser. Indeed, a web browser cannot access local files without an explicit request from the user. This request is possible through different steps in the two versions of TreePics we propose:

1. The web version available at <https://treepics.mnhn.fr> builds a tree in two steps (Fig. 1): first, the pictures that will eventually be displayed are selected, and second, the Newick tree is uploaded. In the http protocol, the web browser does not allow a deferred use of local resources (the pictures) and the pictures should be embedded in the web browser when they are selected. To do that, they are converted in base64-encoded images by the web browser and directly associated to the SVG tree. This methodology reduces the process to only two steps, but the memory used (number of pictures  $\times$  size of each picture) can limit the number of pictures selected. We translated TreePics into two languages (French and English), automatically detected from the language preference by the web browser; it is also possible to switch from one language to the other.

2. A full standalone version (downloadable at <https://treepics.mnhn.fr>) can be installed locally. The process is here divided into three steps: in addition and before the two steps of the web version, the user needs to provide the path to the folder that contains the pictures. As in the web version, the operating system, and not the web browser, will manage the local resources. However, the web browser obtains the data stream (base64) or a list of file names, but not the local path, which needs to be user-specified. This adds a step to the process compared to the web version, but the standalone version does not necessitate an internet connection and reduces the memory needed. The standalone TreePics also exists in French and English and the language is requested when downloading.

### **Data input and visualization**

Two types of files are provided: the pictures, in JPEG or PNG format, and the tree, in Newick format ([https://en.wikipedia.org/wiki/Newick\\_format](https://en.wikipedia.org/wiki/Newick_format)). Values for the tree nodes (typically, statistical support) provided in the Newick tree will be displayed by TreePics. An example (a tree with the associated pictures) is provided in the supplementary material 1.

To associate the pictures to the names of the taxa in the tree, TreePics will compare the names of the pictures and the names of the taxa in the tree. The rule is simple: the name of the picture should be included entirely (without the extension) in the name of the taxon in the tree. The name of the picture should contain only alphanumeric characters, dashes (“-”) and underscores (“\_”); in addition to the constraints linked to the OS (e.g., no `~@#$$%^&*()` for UNIX), it should not contain an exclamation point (!) or quotes (‘), as these are used by TreePics as internal delimiters.

When a picture is associated to a name in the tree, the part of the name corresponding to the name of the picture appears in blue. Pictures are visualized as thumbnails at the tip of each branch, and when the mouse rolls over a thumbnail, a pop-up window opens with a zoom of the picture. A click on a thumbnail selects it and places it in a basket below the tree. The pictures placed in the basket can be visualized in a larger format in a separate window for easy comparison (Fig. 2). Pictures can be unselected either by clicking on the thumbnail or by clicking on the red cross in the basket. A click on the zoom icon on the picture will open it in a new tab on full screen. Finally, clicking on the “print” button will save the tree with the thumbnails in pdf format.

TreePics

[Contact](#)
[Download](#)
[Feedback](#)

---

### Associating images to the tips of a phylogenetic tree

**Rationale**

Several web applications and programs can be used to visualize and edit trees in the **Newick** format (e.g., [FigTree](#)). However, to our knowledge, they do not offer the possibility to easily and automatically associate pictures to the branch tips of a tree. The standalone software **TreePics** has been developed to meet this need. To avoid problems linked to the transfer of pictures from a client to a server, all operations take place on the client in the web browser. **TreePics** is a web application (HTML5, JavaScript, CSS3) that does not need to be run on a web server: if downloaded locally, it can be used without an internet connection.

**Limitations**

- The web browser will allow selection of multiple pictures if the web browser is HTML5 compatible. The most recent versions of all web browsers are compatible with HTML5.
- When using **TreePics** with Google Chrome on a Windows operating system, picture file names are limited and only ~ 2000 files can be selected at a time. There are no such restrictions with Firefox.
- Depending on the RAM available, the number of thumbnails to be displayed and the size of the pictures, the web browser may freeze if it needs more memory than allocated.

**How to cite **TreePics****

Pullandre N., Bisset J., Chagnoux S., Caviller E. **TreePics**: visualizing trees with pictures. Submitted to European Journal of Taxonomy

### A tree in three steps

**1 - Select the folder that contains the pictures**

All operations take place on the client and the pictures are not uploaded on the server. It is thus necessary to give the path to the folder that contains the pictures. The folder name is stocked and will be displayed again at the next use of **TreePics**. Pressing "Enter" or "Tab" will validate this step.

Windows, e.g.: C:\Programme\MolecularEvolution\Images  
Linux or Mac, e.g.: /apps/Projet/MolecularEvolution/Images

**2 - Selecting the pictures (.jpg, .png)**

Please select the pictures that will eventually be associated with the branch tips. You are allowed to select as many pictures as desired – only files with names partially matching the tip labels of your phylogenetic tree will be displayed. There is no uploading: only the names of the pictures are needed.

**3 - Selecting the Newick tree**

Please select your phylogenetic tree in Newick format. (Convert your tree in Newick format using e.g. [FigTree](#) if needed). The characters “” and “|” in the tip labels will not be taken into account.

**Viewing the tree**

You can expand thumbnails by rolling your mouse over the image.  
To view multiple images side by side, select images to place into the basket. Thumbnails are selected and unselected by directly clicking on the image. Images in the basket can be viewed in a larger format in a separate window.  
The tree can also be printed. (Printing works best with Chrome)

Remove	Specimen data	Name of the picture	Thumbnail
✘	..im_2007-41024_ophiotoma_acuta	IM_2007-41024.jpg	
✘	..im_2007-41004_ophiotoma_acuta	IM_2007-41004.jpg	
✘	..im_2007-41133_ophiotoma_picturata	IM_2007-41133.jpg	
✘	..im_2007-41182_ophiotoma_jickelli	IM_2007-41182.jpg	
✘	..im_2007-41144_ophiotoma_jickelli	IM_2007-41144.jpg	

**Fig. 1.** Screenshot of the page with tree visualizations and associated images, showing the phylogenetic trees and the associated thumbnails. The selected thumbnails (red squares) also appear in the list below the tree.

## Restrictions

Because TreePics uses a web browser, the problems linked to the different operating systems are minimized, but not totally removed. In particular, the classic rules for naming the files should be followed (e.g., no space on Mac). We tested several web browsers: while Google Chrome makes better use of the memory than Firefox, it also limits the number of files that can be selected (according to their size) compared to Firefox. Using a PC with 16GB RAM and Firefox, we were able to visualize a tree with 3000 pictures. In any case, the full standalone version is more robust and should be preferred for large datasets.

## Data accessibility

The web version can be accessed at <https://treepics.mnhn.fr>. The full standalone version can be downloaded on the same page.

## Supplementary material

A test dataset, with a tree and the associated pictures ([supplementary material](#)).

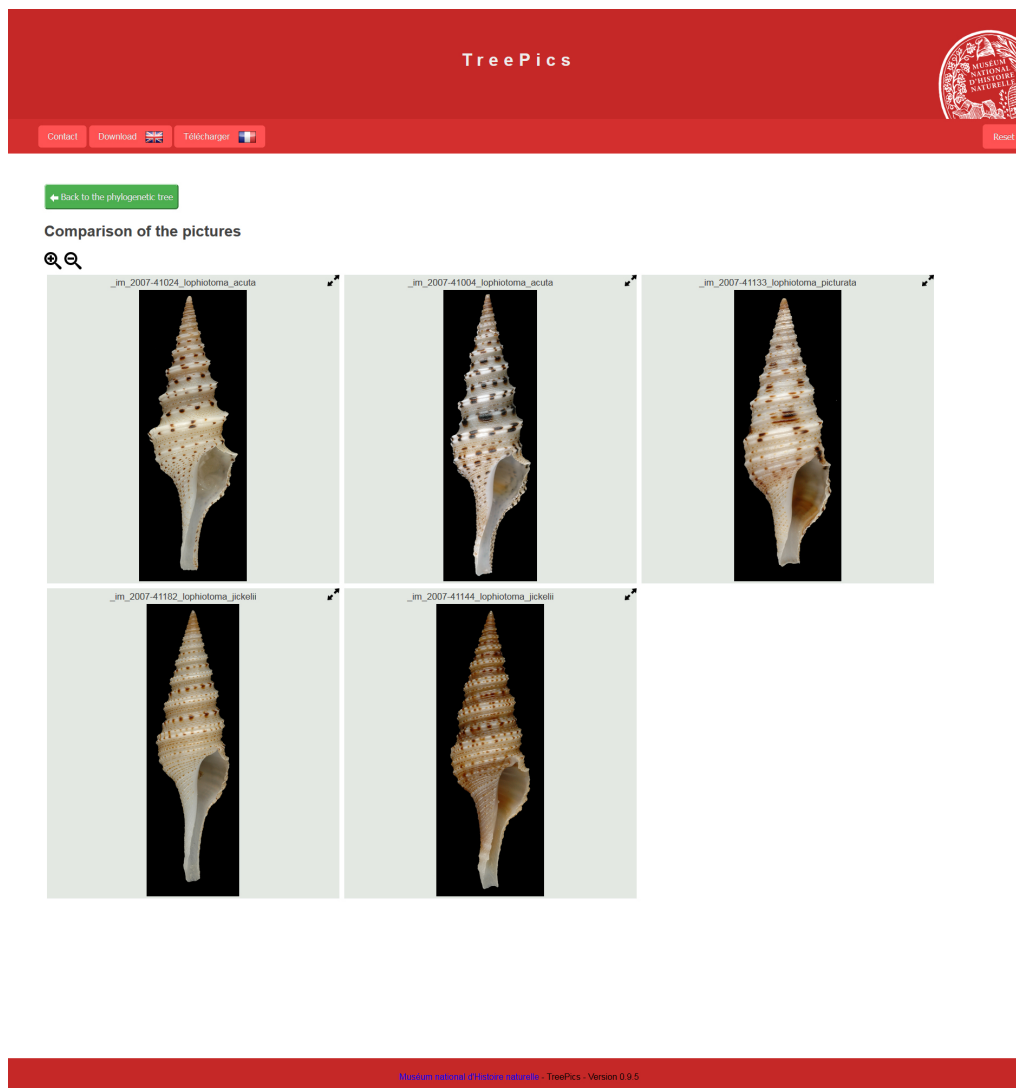


Fig. 2. Screenshot of the page with the specimens selected on the previous page.

## Acknowledgments

TreePics is the result of discussions with the people working with the INVMAR MNHN collections. The authors thank all the people who produced and analyzed the huge amount of data DNA sequences and pictures recorded in the MNHN databases, from which the need for TreePics emerged. In particular, we thank the 3E team (Exploration, Espèces Evolution) of the ISYEB, the INVMAR team that manages the collections, the SSM (Service de Systématique Moléculaire, UMS 2700 CNRS-MNHN) where most of the DNA data used to test TreePics were produced, and the DSI (Division des Systèmes d'Information) of the MNHN. We also thank the team of testers who tracked the bugs in TreePics, and Mark Phuong for his help in correcting the English. This project benefitted from the CONOTAX project, funded by the French ANR (ANR-13-JSV7-0013-01).

## References

- Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- Hebert P.D.N., Cywinska A., Ball S.L. & De Waard J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B* 270: 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Huerta-Cepas J., Serra F. & Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution* 33: 1635–1638. <https://doi.org/10.1093/molbev/msw046>
- Page R.D. 2002. Visualizing phylogenetic trees using TreeView. In: Bateman A., Draghici S., Khurana E., Orchard S. & Pearson W.R. (eds) *Current Protocols in Bioinformatics*: 6.2.1--6.2.15. John Wiley and Sons, New York. <https://doi.org/10.1002/0471250953.bi0602s01>
- Puillandre N., Bouchet P., Boisselier-Dubayle M.-C., Brisset J., Buge B., Castelin M., Chagnoux S., Christophe T., Corbari L., Lambourdière J., Lozouet P, Marani G., Rivasseau A., Silva N., Terryn S., Tillier S., Utge J. & Samadi S. 2012. New taxonomy and old collections: integrating DNA barcoding into the collection curation process. *Molecular Ecology Resources* 12: 396–402. <https://doi.org/10.1111/j.1755-0998.2011.03105.x>
- Rambaut A. & Drummond A. 2009. FigTree v1. 3.1. Available from <http://tree.bio.ed.ac.uk/software/figtree/> [accessed 19 Sep. 2017].
- Sangster G. 2009. Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. *Proceedings of the Royal Society B* 276: 3185–3191. <https://doi.org/10.1098/rspb.2009.0582>
- Schmidt T., Hummel S. & Herrmann B. 1995. Evidence of contamination in PCR laboratory disposables. *Naturwissenschaften* 82: 423–431. <https://doi.org/10.1007/BF01133677>
- Smits S.A. & Ouverney C.C. 2010. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PloS One* 5: e12267. <https://doi.org/10.1371/journal.pone.0012267>
- Zuccon A. & Zuccon D. 2010. MrEnt v. 2.2. Computer program distributed by the authors.

*Manuscript received: 9 December 2016*

*Manuscript accepted: 9 January 2017*

*Published on: 26 September 2017*

*Topic editor: Koen Martens*

*Desk editor: Kristiaan Hoedemakers*

Printed versions of all papers are also deposited in the libraries of the institutes that are members of the *EJT* consortium: Muséum national d'Histoire naturelle, Paris, France; Botanic Garden Meise, Belgium; Royal Museum for Central Africa, Tervuren, Belgium; Natural History Museum, London, United Kingdom; Royal Belgian Institute of Natural Sciences, Brussels, Belgium; Natural History Museum of Denmark, Copenhagen, Denmark; Naturalis Biodiversity Center, Leiden, the Netherlands; Museo Nacional de Ciencias Naturales-CSIC, Madrid, Spain; Real Jardín Botánico de Madrid CSIC, Spain.