SOFTWARE   **Open Access**

PastView: a user-friendly interface to explore ancestral scenarios

François Chevenet[1,2*] , Guillaume Castel[3], Emmanuelle Jousselin[3] and Olivier Gascuel[2,4]

Abstract

Background: Ancestral character states computed from the combination of phylogenetic trees with extrinsic traits are used to decipher evolutionary scenarios in various research fields such as phylogeography, epidemiology, and ecology. Despite the existence of powerful methods and software in ancestral character state inference, difficulties may arise when interpreting the outputs of such inferences. The growing complexity of data (trees, annotations), the diversity of optimization criteria for computing trees and ancestral character states, the combinatorial explosion of potential evolutionary scenarios if some ancestral characters states do not stand out clearly from others, requires the design of new methods to explore associations of phylogenetic trees with extrinsic traits, to ease the visualization and interpretation of evolutionary scenarios.

Result: We developed PastView, a user-friendly interface that includes numerical and graphical features to help users to import and/or compute ancestral character states from discrete variables and extract ancestral scenarios as sets of successive transitions of character states from the tree root to its leaves. PastView provides summarized views such as transition maps and integrates comparative tools to highlight agreements or discrepancies between methods of ancestral annotations inference.

Conclusion: The main contribution of PastView is to assemble known numerical and graphical methods into a multi-maps graphical user interface dedicated to the computing, searching and viewing of evolutionary scenarios based on phylogenetic trees and ancestral character states. PastView is available publicly as a standalone software on www.pastview.org.

Keywords: Phylogeny, Ancestral character states, Evolutionary scenario

Background

Phylogenetic trees are often combined with extrinsic traits to reconstruct evolutionary scenarios in various research fields, including phylogeography, molecular epidemiology, and ecology. The analysis starts with the computation of ancestral annotations from extrinsic traits (discrete variables such as geographic origin, resistance to a treatment, and life history traits), associated with the sampled sequences used to build the phylogenetic tree. It continues with the study of the evolutionary changes from the tree root to its leaves to characterize evolutionary scenarios such as the spread of a disease, the dynamic of a drug resistance, or shifts in ecological habitats.

There are numerous methods and software applications in ancestral character state inferences, among the most used are SIMMAP [1], BEAST [2], SpreaD3 [3], and various functions implemented in R packages [4, 5]. However, many difficulties may arise when interpreting the outputs of these methods. First, there is a diversity of optimization criteria for computing trees and ancestral annotations (parsimony, maximum likelihood, Bayesian inference) and each of them involves various models of evolution; these methods may all yield different results that need to be compared. Second, the tree size and the complexity of the annotations can be an inconvenience with respect to computation time, and the results often require tedious graphical analyses to be interpreted. Third, although probabilistic models give more accurate results than other methods, if some ancestral characters do not stand out clearly from others (i.e. if they are not much more likely), these methods

* Correspondence: francois.chevenet@ird.fr
[1]MIVEGEC, Université de Montpellier, CNRS, IRD, Montpellier, France
[2]LIRMM, Université de Montpellier, CNRS, Montpellier, France
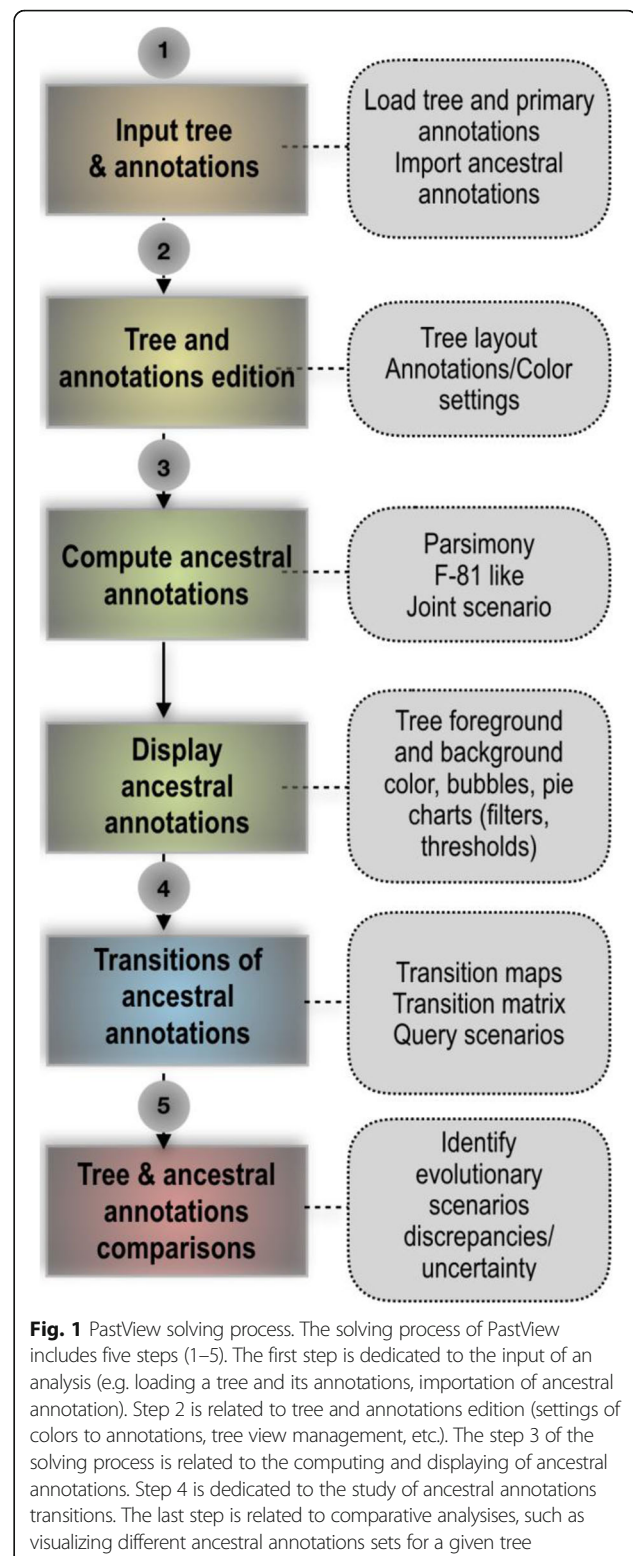Full list of author information is available at the end of the article

may produce a combinatorial explosion of potential evolutionary scenarios.

To the best of our knowledge, there are no tools available to compare different sets of ancestral annotations and to find common patterns across multiple evolutionary scenarios from a beam of ancestral annotation transitions. We thus present PastView that provides numerical and graphical tools to facilitate the interpretation of evolutionary scenarios derived from phylogenetic trees with ancestral character states.
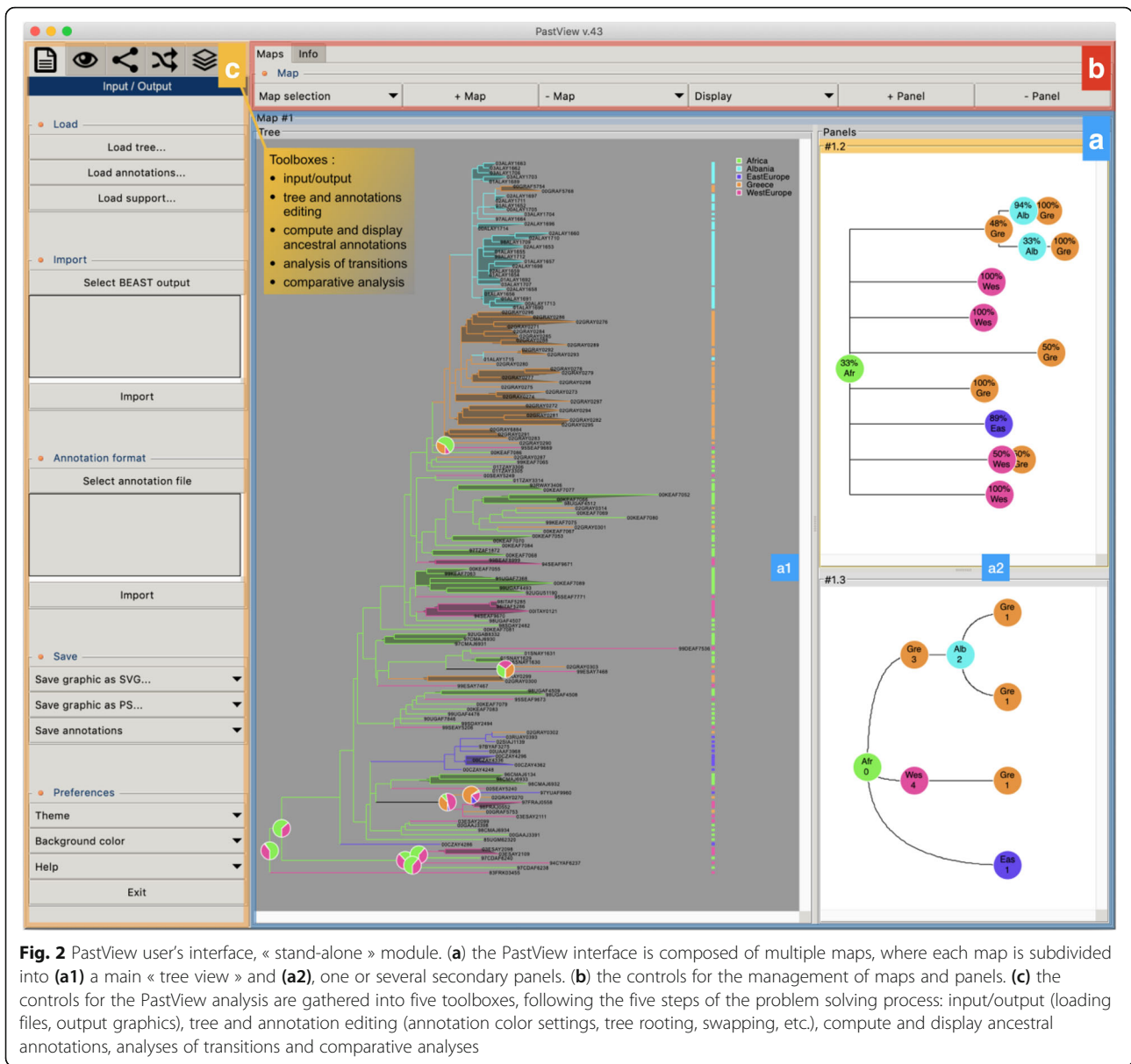
Implementation

PastView is written in Tcl/Tk. It is an open source, cross-platform, standalone editor available for Windows and Unix-like systems including OSX. The PastView process includes five steps (Fig. 1): (1) data input (tree and annotations), (2) tree and/or annotations edition, (3) computing and displaying of ancestral annotations, (4) study of ancestral annotations transitions, and (5) comparative analysis between different methods of ancestral annotation inference. The PastView user interface/architecture (Fig. 2) of PastView is based on multi-maps, where each map is subdivided into a main view for the tree display and secondary view(s). PastView's features are grouped into five toolboxes (Fig. 2c), following its five main steps:

- The "Input/output" toolbox includes controls for I/O management: loading a phylogenetic tree (Newick format) and ancestral annotations or primary annotations (i.e. annotations for leaves only, see below for computing ancestral annotations) in CSV format. The user can also import data in the NEXUS format, which merges tree topology with ancestral annotations (e.g. BEAST output). PastView's graphical output format is PostScript and/or SVG for high quality graphics and CSV for annotations.
- The "Editing trees and annotations" toolbox includes controls for basic editing of trees and annotations: tree layouts, rooting, swapping, ladderizing, zooming in/out, etc.
- The third toolbox is related to the "Computation and display of ancestral annotations". PastView computes ancestral annotations using parsimony: DELTRAN and ACCTRAN [6–8], and maximum likelihood (ML) assuming an F81-like model of character evolution. The branch lengths (typically obtained from sequences) are rescaled to fit the rate of evolution of the studied character. A constant scaling factor is applied to all branch lengths, with value inferred by ML. PastView provides both the marginal posterior state probabilities (using standard tree traversals [9, 10]) and the joint reconstruction

Fig. 1 PastView solving process. The solving process of PastView includes five steps (1–5). The first step is dedicated to the input of an analysis (e.g. loading a tree and its annotations, importation of ancestral annotation). Step 2 is related to tree and annotations edition (settings of colors to annotations, tree view management, etc.). The step 3 of the solving process is related to the computing and displaying of ancestral annotations. Step 4 is dedicated to the study of ancestral annotations transitions. The last step is related to comparative analysises, such as visualizing different ancestral annotations sets for a given tree

with maximum likelihood (using dynamic programming [11]). Details of ML computations can be found in [12]. PastView can also input ancestral annotations computed using other programs (e.g.

Fig. 2 PastView user's interface, « stand-alone » module. (**a**) the PastView interface is composed of multiple maps, where each map is subdivided into (**a1**) a main « tree view » and (**a2**), one or several secondary panels. (**b**) the controls for the management of maps and panels. (**c**) the controls for the PastView analysis are gathered into five toolboxes, following the five steps of the problem solving process: input/output (loading files, output graphics), tree and annotation editing (annotation color settings, tree rooting, swapping, etc.), compute and display ancestral annotations, analyses of transitions and comparative analyses

BEAST). Ancestral annotations are displayed with color-coded bubbles and pie charts with the use of thresholds/filters to simplify views (e.g. highlighting contentious nodes with several annotations having high but close probabilities, assemble annotations with small probabilities, etc.). Ancestral annotation can also be displayed with a tree foreground color-coded under the constraints of a probability threshold and/or a background color-coded according to the Size criterion (Sz, i.e. number of descendant with the same annotation, as in the PhyloType method [13]).

- Toolbox for analysis of ancestral annotation transitions: a transition is a change in ancestral annotation between subsequent nodes of a rooted

phylogenic tree (Top-Down reading). PastView includes numerical and graphical methods to study transition suites. A summary view of all transitions observed in the tree is called a transition map (Fig. 3). There are three kinds of transition maps included in PastView. With the transition map of type 1, we start with a node corresponding to the ancestral root state i, and then create a node j when a transition i - > j is observed; this procedure is continued in a top-down fashion until we reach the tree tips. This map thus stores all transitions occurring in the tree. In the second type of maps we cluster together identical transitions occurring at different positions and depths in the phylogenetic tree, when they have the same ancestor in the transition
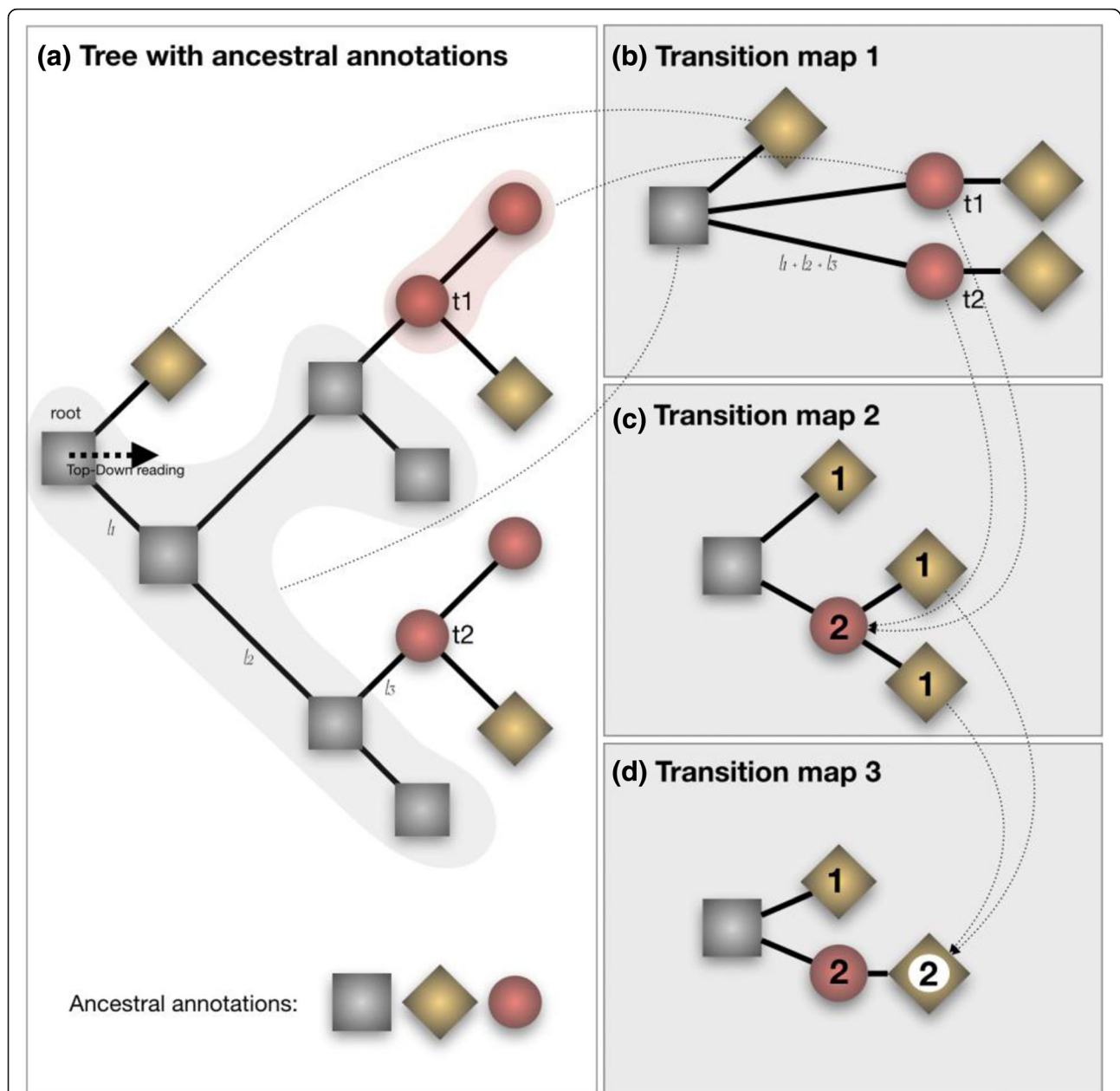
Fig. 3 Tree-like representations of transitions. (**a**) phylogenetic tree with ancestral annotations (gray squares, red circles and gold diamonds). (**b**), (**c**) and (**d**): corresponding transition maps. A transition is a change of an ancestral annotation along a phylogenetic tree (from the root to the leaves). The transition map layout 1 (**b**) is a multifurcating tree-like representation of transitions. A node is created in the map foreach transition observed in the phylogenetic tree (with the exception of the root). It can take into account branch lengths of the tree (sum of branch lengths along the path). The transition map layout 2 (**c**) gather transitions when they share the same father node in the layout 1. The number of transitions collapsed is then displayed (1 from « gray square » or « red circle » to « gold diamond », 2 from « gray square » to « red circle »). The same collapsing process is applied in turn to layout 2 to produce the transition map layout 3. Transition maps can be computed for different sets of ancestral annotations (e.g. different inference methods). When the output consist into probability distributions (e.g. Bayesian or F81-like methods), only the annotation having the greatest probability value is considered

map of type 1. For instance, in the example Fig. 3, the transition from « gray square » to « red circle » is observed twice in layout 1 (nodes t1 and t2), they are then collapsed into one node in layout 2. Lastly, transition maps of type 3 bring together identical

transitions occurring at different positions and depths in the tree, when they have the same ancestral annotation in the transition map of type 2. Transitions maps can be displayed with different layouts (rectangular, slanted, and radial) and can be

drawn dynamically to show the transitions depending on the node distance to the tree root (transition map of type 1 only). Conversely, pointing at a node on a transition map automatically highlights the corresponding part(s) of the phylogenetic tree. The transition maps can be saved in Newick format. The transition matrices provide other overviews of transitions, crossing annotations and giving counts or more elaborated indices of transitions between pairs of annotations (for instance, see [14]). PastView includes two types of matrices. The first one computes the number of transitions in the tree, from one ancestral annotation to another. The second one is a fast, count-based estimation of the relative transition rates, where raw counts are normalized and divided by state priors. PastView offers a query system to search for a given evolutionary path: the user selects a transition sequence (wild cards may be used) and matching tree pathways are identified and

highlighted. For example, the scenario "A * B" lights up the tree pathways entailing the transitions from "A" to "B", independently of the intermediaries (example in Fig. 4e).

- Comparative analyses: this toolbox is dedicated to comparing multiple datasets (trees/ancestral annotations). For example, given a tree and ancestral annotations resulting from different inferences, PastView automatically adjusts tree foreground colors for nodes with similar results, but highlights nodes (bubbles, pie charts) exhibiting incongruences in ancestral annotations (example in Fig. 5a).

Results

In the following, we present two published examples of virus phylogeographic studies re-examined with Past-View. The first one is related to the HIV-1A epidemiological history in Albania. The second is related to the global phylogeography of Dengue serotype 1 virus.
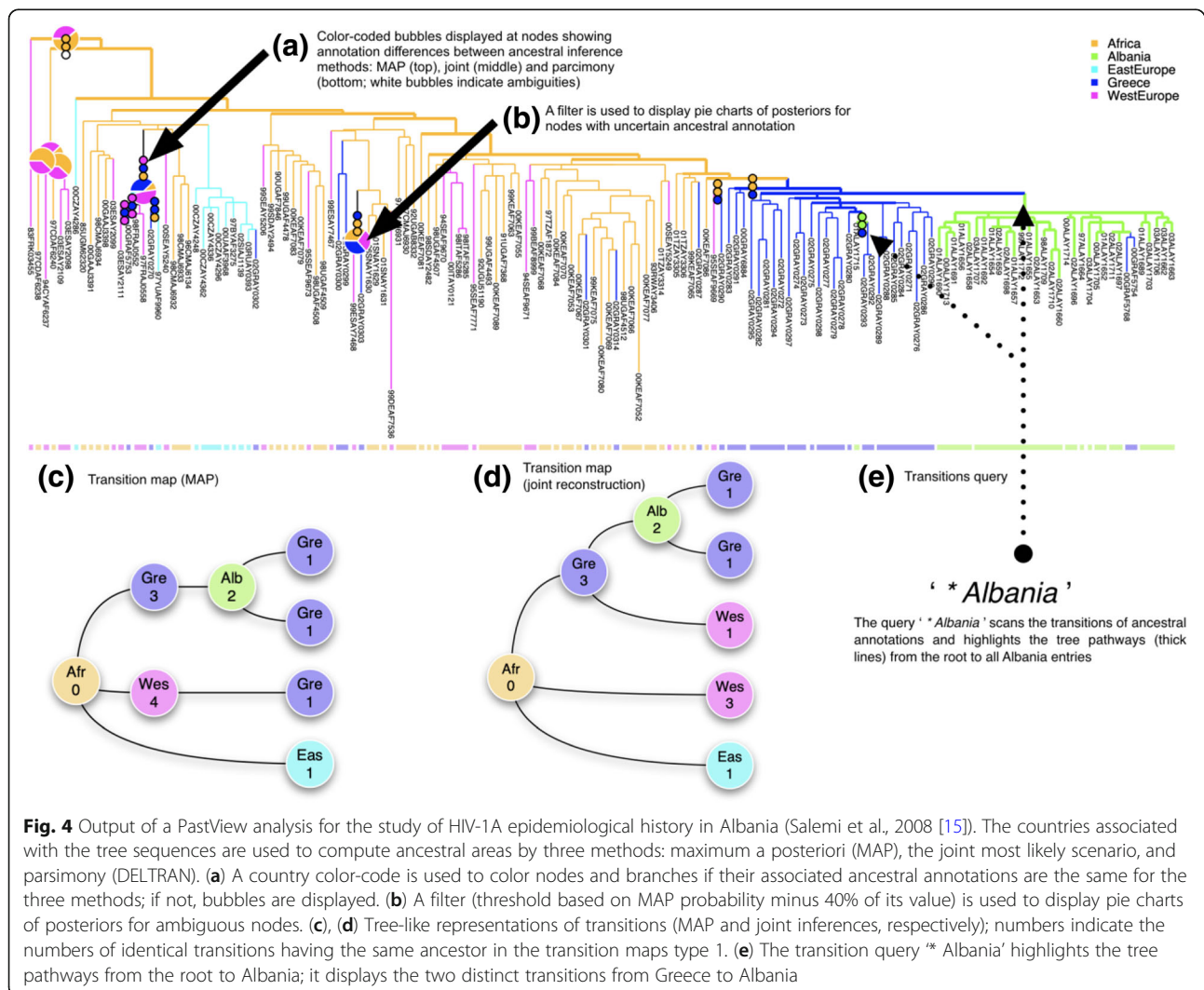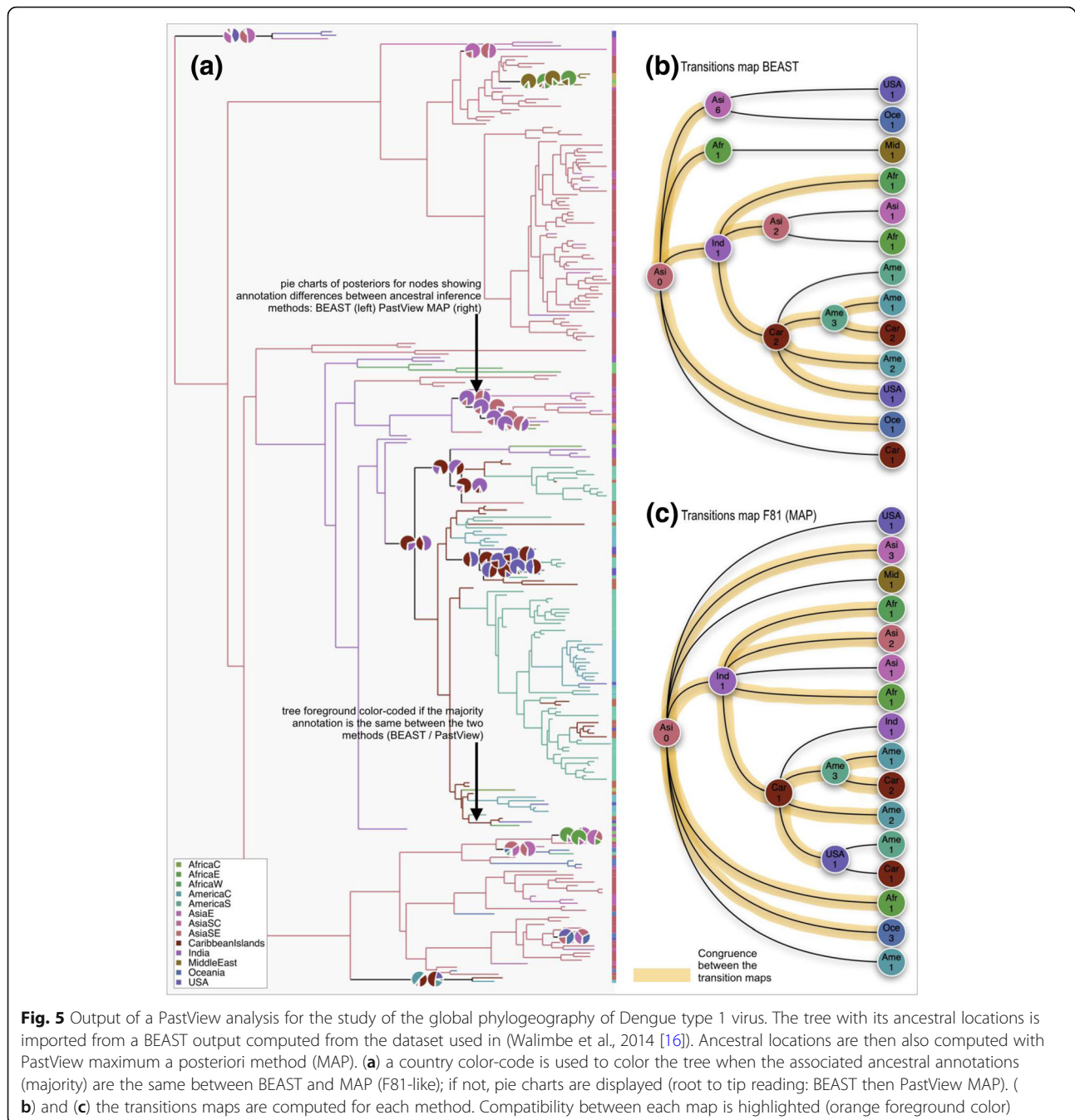

Fig. 4 Output of a PastView analysis for the study of HIV-1A epidemiological history in Albania (Salemi et al., 2008 [15]). The countries associated with the tree sequences are used to compute ancestral areas by three methods: maximum a posteriori (MAP), the joint most likely scenario, and parsimony (DELTRAN). (**a**) A country color-code is used to color nodes and branches if their associated ancestral annotations are the same for the three methods; if not, bubbles are displayed. (**b**) A filter (threshold based on MAP probability minus 40% of its value) is used to display pie charts of posteriors for ambiguous nodes. (**c**), (**d**) Tree-like representations of transitions (MAP and joint inferences, respectively); numbers indicate the numbers of identical transitions having the same ancestor in the transition maps type 1. (**e**) The transition query '* Albania' highlights the tree pathways from the root to Albania; it displays the two distinct transitions from Greece to Albania

Fig. 5 Output of a PastView analysis for the study of the global phylogeography of Dengue type 1 virus. The tree with its ancestral locations is imported from a BEAST output computed from the dataset used in (Walimbe et al., 2014 [16]). Ancestral locations are then also computed with PastView maximum a posteriori method (MAP). (**a**) a country color-code is used to color the tree when the associated ancestral annotations (majority) are the same between BEAST and MAP (F81-like); if not, pie charts are displayed (root to tip reading: BEAST then PastView MAP). (**b**) and (**c**) the transitions maps are computed for each method. Compatibility between each map is highlighted (orange foreground color)

Example 1

Study of HIV-1A epidemiological history in Albania.

Figure 4 shows a simple example of PastView output, in the study of HIV-1A epidemiological history in Albania. The dataset, tree, and origin of the sequences (locations) are from [15] (see also [13]). Ancestral locations are computed by parsimony (DELTRAN), maximum of the marginal posterior probabilities (MAP, F-81 like), and joint most likely scenario. Computation time of ancestral annotations for this small dataset (153 strains) is ~ 2 s with a 3.1 GHz Intel I7 computer. The tree is foreground color-coded if the ancestral annotations are identical between MAP F-81 like, joint most likely reconstruction and parsimony. If not, color-coded bubbles are displayed (Fig. 4a) (in the same order, reading from root to tips). Here, we identify nine nodes with discrepancies, and only four nodes if we compare MAP and joint inferences only. The pie charts display (Fig. 4b) the posteriors for nodes having two or more annotations with probability higher than their MAP F81 like

Chevenet *et al. BMC Evolutionary Biology*   (2019) 19:163

Page 7 of 8

probability value minus 40%. Based on MAP and joint inferences, the transition maps (Fig. 4c and d) are slightly different but output the same global scenario: the virus spread from Africa to West Europe, East Europe and Greece, and then from Greece to Albania, with a few Greek sequences coming back from Albania. This reaches the same general conclusion as [15]: there has been a single major introduction of HIV-1A from Greece followed by a local epidemic spread. This result (Fig. 4e) is highlighted by the thick pathways from the tree root to all entries to Albania.

Example 2

Study of the global phylogeography of dengue serotype 1 virus.

The second example is related to the study of the worldwide phylogeography of Dengue virus of serotype 1 (DENV-1). The dataset, tree (269 strains), and origin of the sequences (13 locations) were obtained from [16]. Tree and ancestral annotations are first computed with BEAST (same parameters / model as [16]), then imported in PastView. Another set of ancestral annotations is computed with PastView (F81-like method, MAP, computation time is ~40s with an 3.1 GHz Intel I7 computer). The tree is then foreground color-coded (Fig. 5a) if the ancestral annotations (MAPs) are the same between the two methods. If not, pie-charts are displayed (from a root to tips reading: left = BEAST, right = PastView). Both analyses reach the same main conclusions as described in [16]: Southeast Asia countries are found to be the most likely origin of the virus dispersion and India played a crucial role in the establishment, evolution and dispersal of the Cosmopolitan (Africa, America, Carribean, East & Southeast Asia) DENV-1 genotype (Fig. 5b and c). The Caribbean region is also found by both methods as the dissemination origin point of the virus to the Americas (South and Central America, then North America). If there is a global consensus for the most ancestral nodes/ transitions, some differences between the results obtained with the two methods exist especially for the most recent nodes/transitions (Fig. 5a, b and c). These disagreements are related mainly to transitions within small clusters of sequences with heterogeneous annotations.

Conclusion

The main contribution of PastView is to assemble a series of numerical and graphical tools into a multi-map interface dedicated to computing, searching and viewing ancestral scenarios. One challenge will be to develop and integrate new methods to compute ancestral annotations combining the advantages of the nuanced outputs of the ML/Bayesian settings, and the overly stringent outputs of the joint and MAP methods, such as the recent decision theory-based PastML approach

[12]. In the long-term, a next step will be to integrate statistical tools and information visualization methods to quickly identify robust evolutionary scenarios, such as a consensus of transition maps between different methods (tree computations, ancestral annotation inferences, see for instance [17]), or for more elaborated transition maps in the case of huge datasets. In the context of massive data, which are more are more common nowadays (e.g. from high-throughput sequencing), it is easy to imagine analyses of trees comprising tens of thousands of tips (e.g. HIV strains) associated to complex, multiple annotations (e.g. clinical data). Then, it will become necessary to design methods that operate on tree topologies / extrinsic traits associations to produce user-friendly views summarizing large, complex evolutionary scenarios.

Availability and requirements

Project name: PastView
  Project home page: http://pastview.org/
  Operating system(s): Platform independent
  Programming language: Tcl/Tk
  Other requirements: ActiveTcl 8.6.8 or higher
  License: GNU GPL
  Any restrictions to use by non-academics: none

Availability of data and materials
PastView source code and data (tree, primary and ancestral annotations) of the two examples used in the manuscript are available on http://pastview.org/ and in the PastView package. PastView and the source code is publicly available on www.pastview.org

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]MIVEGEC, Université de Montpellier, CNRS, IRD, Montpellier, France. [2]LIRMM, Université de Montpellier, CNRS, Montpellier, France. [3]CBGP, INRA, CIRAD,

IRD, Montpellier SupAgro, Université de Montpellier, Montpellier, France.
[4]Unité de Bioinformatique Evolutive, C3BI, USR 3756, Institut Pasteur & CNRS, Paris, France.

References

1. Bollback JP. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics. 2006;7:88.
2. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1). https://doi.org/10.1093/ve/vey016. eCollection 2018 Jan.
3. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. Sprea3D: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol Biol Evol. 2016;33(8):2167–9.
4. Paradis E, Claude J, Strimmer K. APE: analyses of Phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–90.
5. Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2011;3:217–23.
6. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool. 1971;20(4):406–16.
7. Farris JS. Methods for computing Wagner trees. Syst Zool. 1970;19(1):83–92.
8. Swofford D, Maddison W. Reconstructing ancestral character states under Wagner parsimony. Math Biosci. 1987;87(2):199–229.
9. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981;17(6):368–76.
10. Gascuel O, Steel M. Predicting the ancestral character changes in a tree is typically easier than predicting the root state. Syst Biol. 2014;63(3):421–35.
11. Pupko T, Pe'er I, Shamir R, Graur D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol and Evo. 2000;17(6):890–6.
12. Ishikawa S, Zhukova A, Iwasaki W, Gascuel O. Fast likelihood method to reconstruct and visualize ancestral scenarios. Mol Biol Evol. 2019. https://doi.org/10.1093/molbev/msz131.
13. Chevenet F, Jung M, Peeters M, de Oliveira T, Gascuel O. Searching for virus phylotypes. Bioinformatics. 2013;29(5):561–70.
14. Maddison D, Maddison W. MacClade 4; 2003.
15. Salemi M, de Oliveira T, Ciccozzi M, Rezza G, Goodenow MM. High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania. PLoS One. 2008. https://doi.org/10.1371/journal.pone.0001390.
16. Walimbe AM, Lotankar M, Cecilia D, Cherian SS. Global phylogeography of dengue type 1 and 2 viruses reveals the role of India. Infect Genet Evol. 2014;22:30–9.
17. Cui Y, Jansson J, Sung WK. Polynomial-time algorithms for building a consensus MUL-tree. J Comput Biol. 2012;19(9):1073–88.

Publisher's Note