Fennec - Functional Exploration of Natural Networks and Ecological Communities

Markus J. Ankenbrand[1], Sonja Hohlfeld[2], Frank Förster[2,3,4], Alexander Keller[1,2,3]

September 27, 2017

[1]Department of Animal Ecology and Tropical Biology, University of Würzburg, Würzburg, Germany
[2]Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany
[3]Department of Bioinformatics, University of Würzburg, Würzburg, Germany
[4]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Applied Ecology and Bioresources, Gießen, Germany

Abstract

Background: Assessment of species composition in ecological communities and networks is an important aspect of biodiversity research. Yet, for many ecological questions the ecological properties (traits) of organisms in a community are more informative than their scientific names. Furthermore, other properties like threat status, invasiveness, or human usage are relevant for many studies, but they can not be directly evaluated from taxonomic names alone. Despite the fact that various public databases collect such trait information, it is still a tedious manual task to enrich existing community tables with those traits, especially for large data sets. For example, nowadays, meta-barcoding or automatic image processing approaches are designed for high-throughput analyses, yielding thousands of taxa for hundreds of samples in very short time frames.

Results: Here we present the Fennec, a web-based workbench that eases this process by mapping publicly available trait data to the user's community tables in an automated process. We applied our novel approach to a case study in pollination ecology to demonstrate the usefulness of the Fennec. The range of topics covered by the case study includes specialization, invasiveness, vulnerability, and agricultural relevance.

Significance: The Fennec is a free web-based tool that simplifies the inclusion of known species traits in ecological community analyses. We encourage scientists

to participate in trait data submission to existing trait databases and to use the Fennec for their analysis. A public instance containing various traits related to pollination ecology is available at http://fennec.molecular.eco.

Introduction

An important task in biodiversity research is the analysis of species composition of ecological communities and networks. This can be done using traditional methods and more recently also with analytical methods designed for large scale sample processing, like DNA (meta-)barcoding (Keller, Danner, et al., 2015) or automated image analysis (Oteros et al., 2015). Such experiments usually yield amounts of data that are hard to cope with manually (e.g. thousands of operational taxonomic units (OTUs) for hundreds of samples). Therefore, general tools for the automated analysis for taxonomic identification from the raw data have been developed, e.g. QIIME (Caporaso et al., 2010), mothur (Schloss et al., 2009), MEGAN (Huson et al., 2016), VAMPS (Huse et al., 2014) and MicrobiomeAnalyst (Dhariwal et al., 2017). However, most of those tools focus only on the taxonomic composition of the communities. Yet, the relevant ecological or socio-economical questions can often not be answered by looking at taxonomic names alone (Junker et al., 2015; Xu et al., 2014).

Metadata for each species (including life-history traits and other features like conservation status, invasiveness, human usage) is required to answer them. In microbial community ecology, the development of tools has already been initiated that aim to automatically map taxonomy information to functional traits. For examples, to predict the functional profile of microbes, the 16S rRNA sequences can be compared against a database with fully sequenced and annotated bacterial genomes (Aßhauer et al., 2015; Keller, Horn, et al., 2014; Langille et al., 2013). To our knowledge, it remains to date a manual effort to enrich eukaryotic communities similarly with trait meta-data, although such information is already publicly available.

There are international efforts to create databases providing trait information for eukaryotes and prokaryotes, e.g. the LEDA Traitbase (Kleyer et al., 2008), the TRY global plant trait database (Kattge et al., 2011), and BacDive for microbial traits (Söhngen et al., 2016), just to mention a few of many here. On the top-level, TraitBank (Parr, Wilson, Schulz, et al., 2014), which is part of the Encyclopedia of Life project (Parr, Wilson, Leary, et al., 2014) aggregates this information from different sources. These sources are of course far from complete, yet the existing data is already highly informative.

But in order to use traits from databases with communities of hundreds or thousands of taxa, tedious manual work is required. To make the most out of it, trait data should be accessible also with automatic batch annotation procedures, not only single manual requests. Furthermore, tools for visualization

and interactive analysis of community data in combination with organismal properties are limited.

Here we present the Fennec, a web-based workbench that helps researchers enrich their taxonomy-based community and interaction tables with relevant traits for their research questions. We integrated basic tools for interactive visualization and analysis of the trait data in context of the communities and networks.

The Fennec is freely available and a public instance is hosted at http://fennec.molecular.eco. This instance currently holds 1.6 million organisms and 207 thousand trait entries gathered from various sources, currently restricted to traits relevant to pollination ecology. User-provided community and network data can be readily mapped to these traits. We aim to extend this set of traits to cover more research areas in the future, but also allow users to contribute traits to the general public database. An alternative option to use the Fennec is to download a docker container containing the software and run it locally, where private data can be used for the enrichment process.

Case study

A case study showing how Fennec can be used to gain insights into pollination ecology and biomonitoring as a proof-of-concept has been performed using data from a large-scale meta-barcoding study Sickel et al., 2015. In this study 384 pollen samples collected by two closely related solitary bee species of the Megachilidae were analyzed using next-generation sequencing, *Osmia bicornis* and *Osmia truncorum* (synonym *Heriades truncorum*L., 1758). One of the bee species, *O. bicornis* is known to be polylectic, while the other, *O. truncorum* is oligolectic (focusing on Asteraceae). Although the data here originates from next-generation sequencing, any community/network data can be used for the workflow independent of the method for data acquisition.

We chose three exemplary topics to be addressed using the Fennec, with the first related to ecological interactions, followed by one concerning bio-monitoring and lastly one focusing on the socio-economic relevance:

1. Are the two bee species showing preferences and differences between each other in growth habit types of visited plants?
   Given the specialization of *O. truncorum* on Asteraceae (mainly forbs and herbs) one could hypothesize that this bee does not collect pollen from shrubs or trees. *O. bicornis* on the other hand collects from many different taxonomic plant groups. Is this reflected by a variety of growth habits or is there a specialization on plants of a specific growth habit, likewise to the other bee species? This hypotheses address the concept of a correlation between functional and taxonomic diversity of the visited plants.

2. How many (and which) invasive species can be found in the samples? Are

there vulnerable species in the samples? Is the amount of invasive and vulnerable species visited similar in all of the samples and by both species? Monitoring the ranges of invasive as well as threatened plant species is an important task in conservation (Darling et al., 2007; Stout, Jane C. et al., 2009). Using pollen data collected by bees, presence of both types can be monitored by mapping conservation relevant traits to the network data. Further, pollination services by the bee species to both types can be identified.

3. Which plants visited by the bees show agricultural relevance to humans and what is their relative amount compared to the remaining plant species? Bees are used commercially to provide pollination services to agriculturally relevant plants (Klatt et al., 2013). Using traits as *agricultural usage* allow to identify how specific the respective bees were in visiting such plants. On the other hand, solitary bees are important agents to ensure the pollination of wild plant species (Garibaldi et al., 2013), and using these traits it can be monitored whether the bees are mainly attracted to mass flowering crops or also visit other plants in agriculturally shaped landscapes.

Material and Methods

Code Implementation and Accessibility

The Fennec is a web application developed in PHP (http://php.net/) using the Symfony framework (https://symfony.com/) with a front-end using JavaScript (ES6) for interactivity. Server side functionality is bundled in modular web services that are called from the front-end via AJAX requests. Layout and interactivity are provided by multiple well established libraries including bootstrap (https://getbootstrap.com/), jQuery-ui (https://jqueryui.com/), react (https://facebook.github.io/react/), lodash (https://lodash.com/), datatables (https://datatables.net/), and plotly.js (https://plot.ly/javascript/). The code quality is ensured using unit tests and strategies for continuous integration. All data is stored in a PostgreSQL (https://www.postgresql.org/) database, which includes taxonomic, trait, citation and further meta-information (Suppl. Fig. 1). Database accession is handled via the doctrine object relational mapper (http://www.doctrine-project.org/). The community and network data provided by users are uploaded and stored in BIOM format (version 1.0) (McDonald et al., 2012) using the biojs-io-biom library (Ankenbrand et al., 2017).

There are three ways to use the Fennec workbench:

1. Public Instance: We have set up a public instance of the Fennec available at http://fennec.molecular.eco. It's database currently hosts trait data related to pollination ecology and is gathered from various sources. This dataset includes also all traits of the case study presented here, yet not

exclusively. The database is subject to constant further extension with more traits, yet our main goal is to maintain high quality of the data available here. User-provided network and community data is private per default, requiring the user to authenticate using a Fennec, GitHub or Google account.

2. Local Instances: All program code is open-source (MIT License) and freely available at the public repository of GitHub: https://github.com/molbiodiv/fennec. Alongside the pure code, there are ready-to-use docker containers available, which include pre-configured instances to be run in a virtualization environment. More information about the docker environment can be found under https://www.docker.com. These local instances can be set up to be accessible only in a specific local network, so that the software can be run in a restrictive way by workgroups or users. These databases can be filled directly with arbitrary trait data not limited to that included in the public instance.

3. Application Programming Interface (API): We also provide an open API that allows third-party programs to make calls to the public instance, or if available also local instances. We currently implemented an R package that makes use of this API. The usage is not limited to this and we encourage software developers to use this API for integration.

Extensive documentation on the code, but also tutorials for users and guides for administrators to host local instances and software developers to use the API are available at both, the GitHub repository and the public instance.

Mapping community and network data to traits

Minimum requirement for using the Fennec is to provide a community or network table, including taxa as rows and samples (communities)/ taxa (networks) as columns. The cells are considered to represent abundances for the respective combinations, but also presence/absence data can be used. Beyond this table, users may also provide own taxonomy data for taxa (to use an alternative to our default NCBI taxonomy (Federhen, 2012)) and meta-data for the samples. These tables may be uploaded separately as tab-separated text files, or combined in BIOM format (Ankenbrand et al., 2017; McDonald et al., 2012) (figure 1). All tables can be managed using the project page of Fennec (figure 3B). Depending on the user input, taxa are mapped using their scientific names or database identifiers like NCBI-taxonomy-ID (Federhen, 2012) or EOL-ID (Parr, Wilson, Leary, et al., 2014).

Traits to be analyzed with this data can be explored and selected via the intuitive web interface, and added as meta-data to the project (figure 3A). All trait data available for this trait and the taxa of interest are automatically linked into the dataset. If multiple values are available for a single trait and organism combination, they are automatically aggregated, i.e. categorical traits are

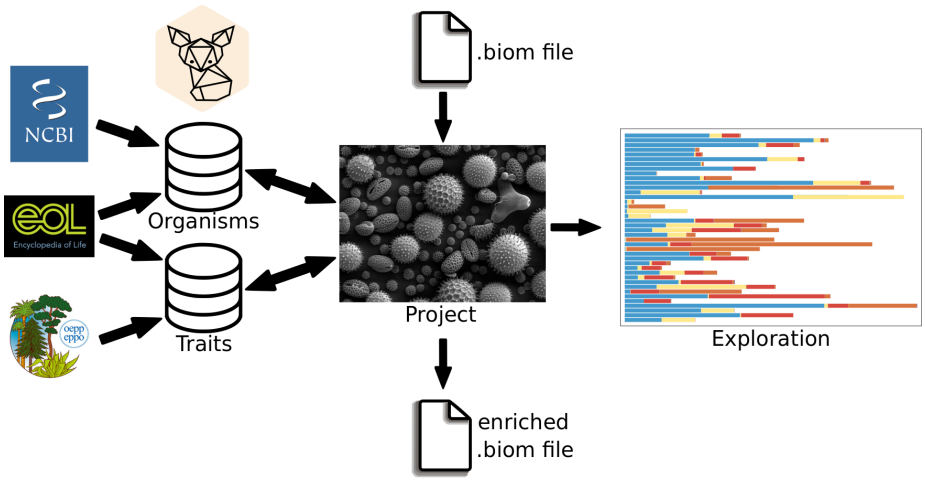

Figure 1: General structure of Fennec. Organisms and traits from different sources like NCBI (Coordinators, 2017), EOL (Parr, Wilson, Leary, et al., 2014), and EPPO are stored in Fennecs database by the administrator. The user imports a community project e.g. in biom format. The organisms in the community are mapped against those in the database and associated traits are used to enrich the metadata. The trait composition can be interactively explored and enriched projects exported e.g. in biom format.

made unique and concatenated and for numerical traits the mean is calculated. To make trait usage as transparent and flexible as possible and to facilitate proper attribution along with the aggregated trait values, trait citations for each individual value are provided alongside the actual traits. Those citations can be exported as a separate table, but are also included as meta-data in any downloaded BIOM file.

After this mapping process, the data is enriched with the selected meta-data and can be further processed with standard analytical and statistical software. For this, the projects can be downloaded as individual tables or again as a single BIOM file that includes all information, which allows fast integration into analysis tools supporting this standard format, e.g. phyloseq (McMurdie et al., 2013) or QIIME (Caporaso et al., 2010).

To provide some basic analytical plots directly in the workbench, we integrated and modified an open-source project for biological data visualization, namely Phinch (Bik et al., 2014). This allows quick interactive exploration of species and trait distributions in each sample, groups, or aggregated by trait types (figure 3D).

Data for the public instance

A public instance of Fennec is hosted at http://fennec.molecular.eco and freely available for direct usage. Taxonomy data in this instance consists of a full representation of the NCBI Taxonomy database (Federhen, 2012 accessed 21-06-2017, >1.6 million taxa). Further a mapping of EOL-IDs (according to http://opendata.eol.org/dataset/hierarchy_entries, accessed on 04/04/2017) has been imported, so that full-text information about taxa is available where EOL offers such (Parr, Wilson, Leary, et al., 2014). Currently and as a starting seed, trait data from TraitBank (Parr, Wilson, Schulz, et al., 2014), EPPO (EPPO, 2017), the World Crops Database (Bijlmakers, 2017), the cavety-nesting bees and wasps database (Budrys et al., 2014, part of the SCALES project (Henle et al., 2014)), and IUCN (IUCN, 2017) have been imported for several plant and bee traits relevant in pollination ecology (table 1), which is subject to continuous extension. We aim to maintain high-quality of these publicly available traits, so that the integration of more traits is a steadily ongoing process. While the bulk of trait data is gathered from databases, in the next release users can also participate in the uploading of trait data, so that this process can be actively supported by the community (see below).

Table 1: Trait types currently imported into the public instance of Fennec. Sources are EPPO (EPPO, 2017), the World Crops Database (Bijlmakers, 2017), TraitBank (Parr, Wilson, Schulz, et al., 2014), SCALES (Budrys et al., 2014; Henle et al., 2014) and IUCN (IUCN, 2017). Along with each type the number of values in the database and the number of distinct organisms with this trait in the database is shown. Invasiveness, conservation status and uses are generally not restricted to plants but the values for those traits as retrieved from TraitBank are.

Trait Type	#values	#organisms	format	source
Plant Traits				
EPPO Categorization	409	409	categorical	EPPO
Invasive In Country	171	171	categorical	TraitBank
Vegetative Spread Rate	1713	1710	categorical	TraitBank
Plant Growth Habit	69781	25186	categorical	TraitBank
Soil Requirements	6019	2102	categorical	TraitBank
Dispersal Vector	686	398	categorical	TraitBank
Flower Color	2808	1916	categorical	TraitBank
Plant Propagation Method	18046	2159	categorical	TraitBank
Life Cycle Habit	21476	18062	categorical	TraitBank
Leaf Color	1838	1835	categorical	TraitBank
Salt Tolerance	1790	1787	categorical	TraitBank
Conservation Status	8431	8247	categorical	TraitBank
Nitrogen Fixation	1852	1849	categorical	TraitBank
Uses	18520	1849	categorical	TraitBank
World Crops Database	508	507	categorical	WCD
Soil pH	3642	1818	numerical	TraitBank
Plant Height	3153	2389	numerical	TraitBank
Leaf Area	504	67	numerical	TraitBank
Bee Traits				
Nest built of	77	76	categorical	SCALES
Foraging mode	92	92	categorical	SCALES
Trophic specialization	86	86	categorical	SCALES
Larval food type	92	91	categorical	SCALES
Landscape type	73	60	categorical	SCALES
Sex ratio (categorical)	60	60	categorical	SCALES
Specialized on	63	63	categorical	SCALES
Nest cells	48	48	numerical	SCALES
Sex ratio	60	60	numerical	SCALES
Body length: female	106	106	numerical	SCALES
Body weight: female	68	68	numerical	SCALES
Generic				

Trait Type	#values	#organisms	format	source
IUCN Red List	44677	44469	categorical	IUCN

Importing organism and trait data

For both, public and local instances, traits are imported using a simple table format containing the organism and trait value, a citation as well as optionally an ontology URL, and source URL as columns, and each entry as a new row. This allows for easy import of traits from various sources. A template for the format for upload is available alongside the Fennec GitHub repository or the public instance.

Two trait formats are currently supported: categorical and numerical. Categorical traits may also include an ontology URL for their value, supporting the hierarchical classification characteristics of some traits. Numerical trait types may be uploaded with an associated unit.

Currently users can upload their own traits as project specific meta-data. It is planned for future releases to upload traits to the public instance using a form. After the submission and verification process, the user may work on this meta-data privately or choose to make it publicly accessible for other researchers. All data marked as public will then undergo a limited manual verification, which e.g. ensures that units are correctly standardized, but does not verify the correctness of the underlying data. Therefore, the user-name of the uploader will be permanently linked to the data to be able to address future changes and updates, and it is required to provide citation information for public records.

For local instances, arbitrary organism and trait data can be imported into the Fennec database by the local administrator using the command line interface (CLI), which are then not subject to a central verification process, but available instantly. This allows for creation of instances tailored to specific organism groups and associated research questions, with the responsibility of the administrator to ensure quality of the imported data. For these public instances, either the same NCBI-taxonomy data can be used or custom taxonomy data provided. Each imported organism receives a unique Fennec-ID which can be linked on-the-fly to other identifiers like NCBI-Taxonomy-IDs or EOL-IDs. The linked EOL-ID is used to provide dynamic content for each organism using the EOL application programming interface (API), where EOL provides such (figure 3C).

Case Study

To demonstrate the analytical potential of the Fennec, we use it to analyse data obtained from pollen collections of the two solitary megachilid bees *Osmia bicornis* and *O. truncorum* in Germany (Sickel et al., 2015). The dataset consists of

384 samples obtained by meta-barcoding using next-generation sequencing of the ITS2 region with the Illumina MiSeq. The data has been downloaded from EBI-SRA project number PRJEB8640 and data preparation as well as taxonomic classification has been performed based on Sickel et al., 2015. The full workflow has been deposited at https://github.com/molbiodiv/meta-barcoding-dual-indexing. This resulted in a table with 1002 plant operational taxonomic units (OTU) and a total count of 6,979,584 observations (sequence reads). For each OTU, the taxonomic lineage and NCBI-taxonomy-ID have been determined during this process by hierarchic taxonomic assignments using UTAX (part of usearch, Edgar, 2010). OTUs with total count of less or equal than 50 across all samples were excluded from the analysis. Samples with less than 10,000 sequence reads remaining have been removed as well. The resulting table consists of 353 plant OTUs corresponding to 216 distinct taxa and 324 samples, which was imported into the Fennec. The total number of reads in this final dataset is 6,663,014. For the plants, the obtained NCBI-taxonomy-ID was used to map the OTUs in the community to organisms in the Fennec database, which resulted in all 353 OTUs being successfully mapped. For the samples, the corresponding bee species were mapped by the scientific name in the meta-data field "beeSpecies".

In the next step, values for the traits listed in Table 1, except "Invasive in" (as this contains only values for USA and samples have been collected in Germany) have been added to the project from the database. Detailed reference information for each individual trait value is given in the supplementary files S2 and S3. This dataset including the traits has then been interactively visualized and analyzed using the built-in modified version of Phinch (Bik et al., 2014) according to the research questions described above. Finally, the enriched dataset has been exported and imported into R (R Core Team, 2017) using shiny-phyloseq (McMurdie et al., 2015) to demonstrate the usability of mapped data in further analyses tools. In particular a DCA ordination has been calculated and visualized with colorization by the trait "Plant Growth Habit". For this purpose OTUs with missing trait values and those with rare variants (keeping only forb/herb, tree, subshrub, shrub/tree, forb/herb/subshrub, forb/herb/vine) were filtered.

Results and Discussion

The Fennec is a useful tool for automated mapping from taxonomic data to functional meta-data of whole communities. This can be done with user-supplied traits or such data-mined from trait databases. A growing public instance is available for analyses in pollination studies. It can be accessed via a graphical web interface and programmatically via an API. Local instances can be used for other and specific traits or organisms. The workbench provides basic visualization options for the mapped data, as well as export options in various file formats to use in downstream analytical software.

Case study

To show the potential of the FENNEC to be used in ecological analysis, we conducted a case study as proof-of-concept for a pollen meta-barcoding data. We address multiple ecological questions and highlights some use cases, where automatic integration of public trait data with the FENNEC has been performed.

Are the two bee species showing preferences and differences between each other in growth habit types of visited plants?

A breakdown of the trait "Plant Growth Habit" for the two bee species separately (visualized via "Donut Partition Chart") reveals that for *O. truncorum* 89% of the taxonomic assignments were mappable to the trait, which resulted in absolute a dominance of "forb/herb" with 87%. This matches our expectations as this bee is specialized on Asteraceae which mostly show this habit.

For *O. bicornis*, 95% of the sequence data was assignable to "Plant Growth Habit", also with "forb/herb" with 65% being the most abundant, but a still considerable amount of 24% as "tree". Likewise to taxonomic specialization, no indication for a specialization on a specific plant growth habit is apparent.

Another interesting observation is the trait coverage when taking abundance into account. While only 85% of OTUs have a value for "Plant Growth Habit", those OTUs contribute 93% to the entire community. Thus the OTUs with missing trait are relatively rare in the community, with the more abundant ones being well-studied in terms of trait data.

Automatically mapped trait data also helps in interpretation of beta-diversity turnover between samples, here collected pollens. For example, ordinations can be visualized with trait data, in our case "Plant Growth Habit", as a split-plot with samples shaped by bee species and plant taxa colored by Plant Growth Habit (figure 2). In our case study, samples are separated as expected by bee species on the first ordination axis with all samples from *O. truncorum* mostly isolated on the right hand side. OTUs localized similarly with possible values for ordination axis 1 were almost exclusively forbs and herbs. The variation of this bee species on the second axis is negligible. For *O. bicornis* there is a substantial spread particularly on the second axis, where plants of type tree seem to concentrate in the upper part. The trait data helps to understand the ecology behind the dataset, indicating plant turnover and eventually also location and landscape changes to be represented on the second axis.

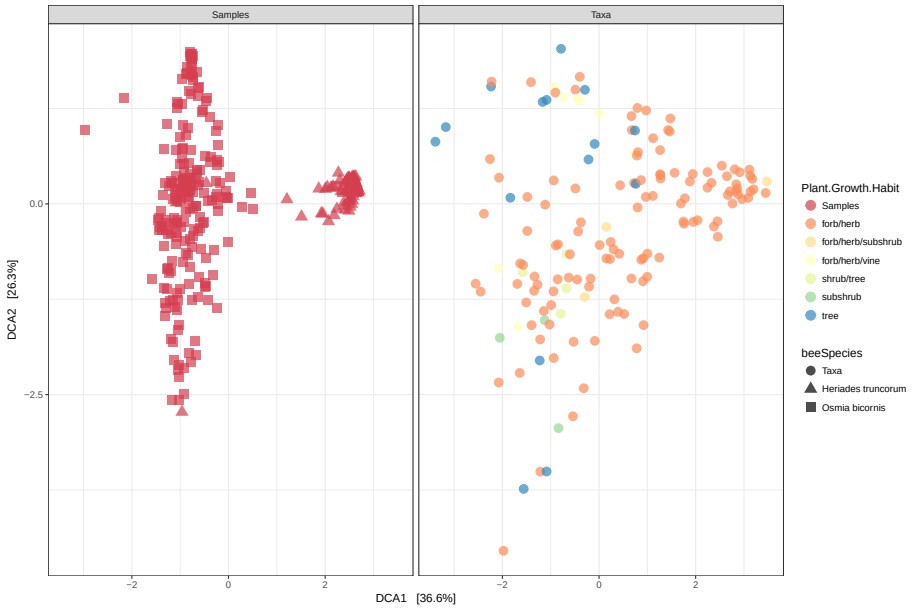

Figure 2: Splitplot of a DCA ordination. Samples are in the left facet with shapes according to by bee species. OTUs are in the right facet with points colored by growth habit (filtered for most common growth habits, species with missing trait have been removed). Samples split nicely by bee on the first axis with *O. truncorum* on the right hand side. The OTUs on the right hand side of the ordination are as expected mainly forb/herb. For *O. bicornis* there is a substantial spread on the second axis.

How many (and which) invasive species can be found in the samples? Are there vulnerable species in the samples? Is the amount of invasive and vulnerable species visited similar in all of the samples and by both species?

The trait "EPPO Categorization" was mapped to our pollen collection data to determine if and to what extend the samples contain species that are regarded as invasive in Europe. One of the visualization methods of the Phinch suite that is integrated into the Fennec, the "Bubble Chart", has been applied to explore this trait. It reveals three samples containing high numbers of invasive species (PoJ74, PoJ236, PoJ244). Further inspection with the integrated meta-data tables showed that PoJ74 and PoJ244 have more than 1000 counts of *Solidago canadensis*, each while PoJ236 has a count of 2779 for *Helianthus tuberosus*.

The trait data is thus sortable regarding abundance of specific traits, referable to organism information but also to samples and their corresponding geographical locations if the data has been collected in such way. It might thus serve as indicator for occurrence of invasive species in geographic regions and used to monitor the spread of invasive species over space and time.

Regarding the occurrence of species with respect to threat status, the pollen data was automatically mapped to the IUCN red list data and the distribution of vulnerable species (as listed by the IUCN) across samples was visualized using the "Bubble Chart", but also a "Taxonomy Bar Chart". These charts illustrate that multiple samples consist almost entirely of "near threatened" species, particularly *Juglans regia*, the english walnut, which experienced strong declines through anthropogenic overuse and lack of replacement plantings. As indicated by the data, it served as a major nutrient source for individual investigated bees.

Which plants visited by the bees are agriculturally relevant to humans and what is their relative amount compared to the remaining plant species?

Finally ecologists (especially in the field of conservation) are often in the difficult situation to somehow quantify economic value of ecosystem services like pollination (Hanley et al., 2015). The Fennec helps in addressing such socio-economic questions by including human usage (as crop) as a trait. All plants listed in the World Crops Database are known to be cultivated by humans for specific purposes (Bijlmakers, 2017). The "Donut Partition Chart" for this trait reveals that 36.7% of plants collected by *O. bicornis* and 7.3% of plants collected by *O. truncorum* are listed in that database (figure 4). This does not yet give more information like the category of crop (e.g. fruits, vegetables, nuts, wood product, etc.) or a real monetary quantification. However this is not a limitation of Fennec but of the underlying data (i.e. if this data is available it can be imported into Fennec and is then automatically available for the community of interest).

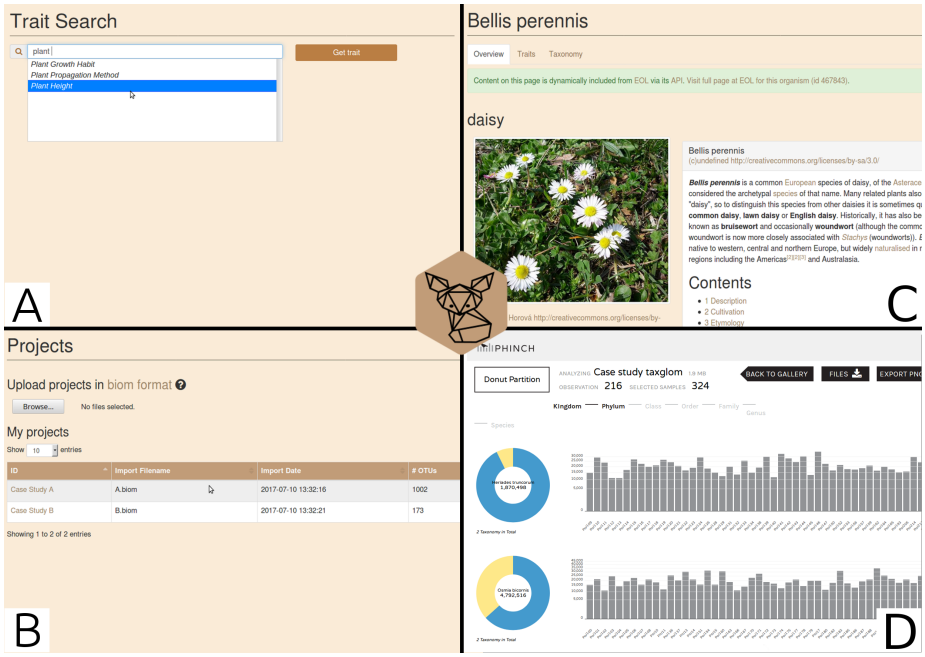

Figure 3: User interface of Fennec. A: Explore all traits which are stored in the database via the trait search. B: Upload community data tables, manage and analyze them. C: Get dynamic content for each organism using the API of EOL. D: Visualize data using the "Donut Partition Chart" of the Phinch suite that is integrated into the Fennec.

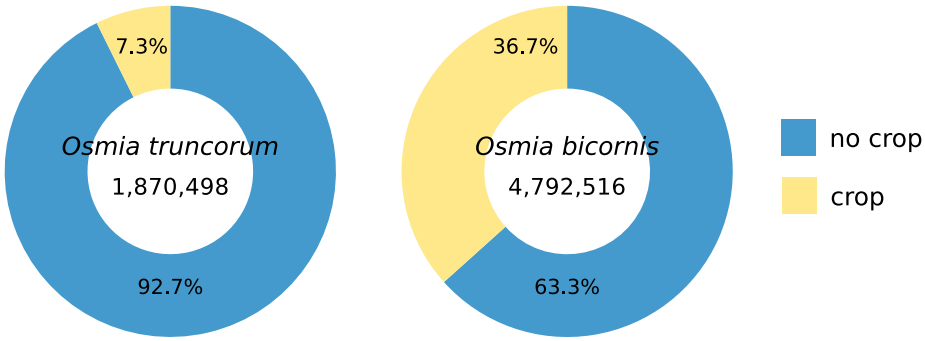

Figure 4: Partition donut charts for the trait "World Crops Database" separated by bee species. Plot has been created with the built in modified version of Phinch.

Outlook and limitations

Fennec has still a number of limitations that will be addressed in future releases. In particular common tasks like filtering and normalization have to be done prior to the upload, as that was not the main focus of the tool. Further, traits for bacteria have proven hard to use for standard 16S microbiome studies as classification can only go down to species or genus level (Werner et al., 2012), while traits may vary on the strain level (Truong et al., 2017). However, one of the main factors restricting Fennecs utility is the currently limited amount of trait data being available in a usable format. One thing that became apparent while building the Fennec is that a lot of trait data is available online but the majority can not be easily used because not adhering to the FAIR principles (Findable, Accessible, Interoperable, Reusable) (Mons et al., 2017; Wilkinson et al., 2016). Licensing of the data is another common problem, data can only be efficiently re-used if it is open and citable in addition to being FAIR (Katz, 2017). An important step for trait data collectors to provide their data in this manner, is to guarantee that the data re-users are able to properly cite all data sources, e.g. by adhering to the FORCE11 data citation principles (Martone, 2014). Fennec supports this by preserving all relevant information. We therefore encourage trait data collectors to make their data available via existing platforms like TraitBank (Parr, Wilson, Schulz, et al., 2014), and with that also usable for downstream analysis tools like Fennec and ultimately to the whole research community.

Conclusion

Fennec as a tool provides valuable assistance to analyze ecological data in the context of organismal information. Both species traits and metadata like threat status and economical importance help to answer different kinds of questions. The public instance can be used as a reference, to try features of Fennec and analyze some datasets with data from other public databases. The possibility to host local instances with own data increases the range of applications. General problems are limited trait data availability, which is however increasing with time and the motivation of more and more scientific journals to make public data deposition mandatory for publications. Beside developing a public automatic mapping procedure, we also aim to demonstrate the importance to make trait data publicly available and how useful it can be in follow up studies. Thus encouraging scientists to submit their data to public databases. Despite those limitations we demonstrated that the Fennec is already able to facilitate ecological community analyses in the light of organismal information. It's usefulness is expected to increase due to continued development guided by user feedback, integration of more analysis tools, better taxonomic resolution, and increasing availability of suitable trait data.

Acknowledgements

MJA was supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Würzburg. We are grateful to the members of the EOL project who were always very quick responding, open and helpful. Many thanks to Anna Voulgari-Kokota and Alice Claßen for beta testing and valuable feedback. We are grateful to Niklas Terhoeven for his suggestions in designing the user interface and Jörg Schultz for constant feedback on the project.

Author Contributions

MJA and AK conceived the project, as well as designed methodology and the software; MJA and SH wrote the code with support by FF; MJA analyzed the case study and drafted the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

Ankenbrand, M. J. et al. (2017). "biojs-io-biom, a BioJS component for handling data in Biological Observation Matrix (BIOM) format". In: *F1000Research* 5, p. 2348. DOI: 10.12688/f1000research.9618.2.

Aßhauer, K. P. et al. (2015). "Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data". In: *Bioinformatics* 31.17, pp. 2882–2884. DOI: 10.1093/bioinformatics/btv287.

Bijlmakers, H. (2017). *World Crops Database - Fruits, vegetables, cereals and other agricultural crops*. World Crops Database. URL: http://world-crops.com/ (visited on 07/25/2017).

Bik, H. M. and P. Interactive (2014). "Phinch: An interactive, exploratory data visualization framework for –Omic datasets". In: *bioRxiv*, p. 009944. DOI: 10.1101/009944.

Budrys, E., A. Budriene, and S. Orlovskyte (2014). *Cavity-nesting wasps and bees database*. URL: http://scales.ckff.si/scaletool/?menu=6&submenu=3 (visited on 09/22/2017).

Caporaso, J. G. et al. (2010). "QIIME allows analysis of high-throughput community sequencing data". In: *Nature Methods* 7.5, pp. 335–336. DOI: 10.1038/nmeth.f.303.

Coordinators, N. R. (2017). "Database Resources of the National Center for Biotechnology Information". In: *Nucleic Acids Research* 45 (D1), pp. D12–D17. DOI: 10.1093/nar/gkw1071.

Darling, J. A. and M. J. Blum (2007). "DNA-based methods for monitoring invasive species: a review and prospectus". In: *Biological Invasions* 9.7, pp. 751–765. DOI: 10.1007/s10530-006-9079-4.

Dhariwal, A. et al. (2017). "MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data". In: *Nucleic Acids Research* 45 (W1), W180–W188. DOI: 10.1093/nar/gkx295.

Edgar, R. C. (2010). "Search and clustering orders of magnitude faster than BLAST". In: *Bioinformatics* 26.19, pp. 2460–2461. DOI: 10.1093/bioinformatics/btq461. eprint: /oup/backfile/content_public/journal/bioinformatics/26/19/10.1093_bioinformatics_btq461/2/btq461.pdf.

EPPO (2017). *EPPO Global Database (available online)*. URL: https://gd.eppo.int/ (visited on 09/22/2017).

Federhen, S. (2012). "The NCBI Taxonomy database". In: *Nucleic Acids Research* 40 (D1), pp. D136–D143. DOI: 10.1093/nar/gkr1178.

Garibaldi, L. A. et al. (2013). "Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance". In: *Science* 339.6127, pp. 1608–1611. DOI: 10.1126/science.1230200. eprint: http://science.sciencemag.org/content/339/6127/1608.full.pdf.

Hanley, N. et al. (2015). "Measuring the economic value of pollination services: Principles, evidence and knowledge gaps". In: *Ecosystem Services* 14, pp. 124–132. DOI: 10.1016/j.ecoser.2014.09.013.

Henle, K. et al. (2014). "Scaling in Ecology and Biodiversity Conservation". In: *Advanced Books* 1, e1169. DOI: 10.3897/ab.e1169.

Huse, S. M. et al. (2014). "VAMPS: a website for visualization and analysis of microbial population structures". In: *BMC Bioinformatics* 15, p. 41. DOI: 10.1186/1471-2105-15-41.

Huson, D. H. et al. (2016). "MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data". In: *PLOS Computational Biology* 12.6, e1004957. DOI: 10.1371/journal.pcbi.1004957.

IUCN (2017). *IUCN Red List of Threatened Species. Version 2017-1*. URL: http://www.iucnredlist.org/ (visited on 09/22/2017).

Junker, R. R., N. Blüthgen, and A. Keller (2015). "Functional and phylogenetic diversity of plant communities differently affect the structure of flower-visitor interactions and reveal convergences in floral traits". In: *Evolutionary Ecology* 29.3, pp. 437–450. DOI: 10.1007/s10682-014-9747-2.

Kattge, J. et al. (2011). "TRY – a global database of plant traits". In: *Global Change Biology* 17.9, pp. 2905–2935. DOI: 10.1111/j.1365-2486.2011.02451.x.

Katz, D. S. (2017). *FAIR is not fair enough*. Daniel S. Katz's blog. URL: https://danielskatzblog.wordpress.com/2017/06/22/fair-is-not-fair-enough/ (visited on 09/13/2017).

Keller, A., N. Danner, et al. (2015). "Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples". In: *Plant Biology* 17.2, pp. 558–566. DOI: 10.1111/plb.12251.

Keller, A., H. Horn, et al. (2014). "Computational integration of genomic traits into 16S rDNA microbiota sequencing studies". In: *Gene* 549.1, pp. 186–191. DOI: 10.1016/j.gene.2014.07.066.

Klatt, B. K. et al. (2013). "Bee pollination improves crop quality, shelf life and commercial value". In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1775. DOI: 10.1098/rspb.2013.2440. eprint: http://rspb.royalsocietypublishing.org/content/281/1775/20132440.full.pdf.

Kleyer, M. et al. (2008). "The LEDA Traitbase: a database of life-history traits of the Northwest European flora". In: *Journal of Ecology* 96.6, pp. 1266–1274. DOI: 10.1111/j.1365-2745.2008.01430.x.

Langille, M. G. I. et al. (2013). "Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences". In: *Nature Biotechnology* 31.9, pp. 814–821. DOI: 10.1038/nbt.2676.

Martone, M. ((2014). *Joint Declaration of Data Citation Principles*. FORCE11. San Diego CA. URL: https://force11.org/datacitation (visited on 09/13/2017).

McDonald, D. et al. (2012). "The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome". In: *GigaScience* 1, p. 7. DOI: 10.1186/2047-217X-1-7.

McMurdie, P. J. and S. Holmes (2013). "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." In: *PloS one* 8.4, e61217. DOI: 10.1371/journal.pone.0061217.

McMurdie, P. J. and S. Holmes (2015). "Shiny-phyloseq: Web application for interactive microbiome analysis with provenance tracking". In: *Bioinformatics (Oxford, England)* 31.2, pp. 282–283. DOI: 10.1093/bioinformatics/btu616.

Mons, B. et al. (2017). "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud". In: *Information Services & Use* 37.1, pp. 49–56. DOI: 10.3233/ISU-170824.

Oteros, J. et al. (2015). "Automatic and Online Pollen Monitoring". In: *International Archives of Allergy and Immunology* 167.3, pp. 158–166. DOI: 10.1159/000436968.

Parr, C. S., N. Wilson, P. Leary, et al. (2014). "The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth". In: *Biodiversity Data Journal* 2, e1079. DOI: 10.3897/BDJ.2.e1079.

Parr, C. S., N. Wilson, K. Schulz, et al. (2014). "TraitBank: Practical semantics for organism attribute data". In: *Semant Web–Interoperability, Usability, Appl an IOS Press J*, pp. 650–1860.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schloss, P. D. et al. (2009). "Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities". In: *Applied and Environmental Microbiology* 75.23, pp. 7537–7541. DOI: 10.1128/AEM.01541-09.

Sickel, W. et al. (2015). "Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach". In: *BMC Ecology* 15, p. 20. DOI: 10.1186/s12898-015-0051-y.

Söhngen, C. et al. (2016). "BacDive – The Bacterial Diversity Metadatabase in 2016". In: *Nucleic Acids Research* 44 (D1), pp. D581–D585. DOI: 10.1093/nar/gkv983.

Stout, Jane C. and Morales, Carolina L. (2009). "Ecological impacts of invasive alien species on bees". In: *Apidologie* 40.3, pp. 388–409. DOI: 10.1051/apido/2009023.

Truong, D. T. et al. (2017). "Microbial strain-level population structure and genetic diversity from metagenomes". In: *Genome Research* 27.4, pp. 626–638. DOI: 10.1101/gr.216242.116.

Werner, J. J. et al. (2012). "Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys". In: *The ISME Journal* 6.1, pp. 94–103. DOI: 10.1038/ismej.2011.82.

Wilkinson, M. D. et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3, sdata201618. DOI: 10.1038/sdata.2016.18.

Xu, Z. et al. (2014). "Which is more important for classifying microbial communities: who's there or what they can do?" In: *The ISME Journal*, pp. 1–3. DOI: 10.1038/ismej.2014.157.