Names

Nomina si nescis,
perit et cognitio rerum

Linnaeus


Quella cosa tonda per scavare cibo…..

Identification

The “name” of an organism is the only access to information about its taxonomy, ecology, etc., which have been collected in centuries of scientific research.

Classification and identification belong to two different operational processes.

Classification is the JOB of taxonomists

BUT

Identification can be FUN for anybody

Identification

Today, however, identification can be achieved not only by using a decision tree, but also by other approaches.

Thus, there exist three main approaches:

1. Decision trees (digital or paper printed)
2. DNA barcode
3. Automated image recognition

While the first approach provides observers with some tool to perform an identification, forcing them to make decisions through the process, the others normally output a name (or an array of names) which could/should match the observed specimen.

Identification

Decision trees

Require **human intervention** at each step of the identification process

Do require a certain **skill level**, depending on the decision tree adopted in the process

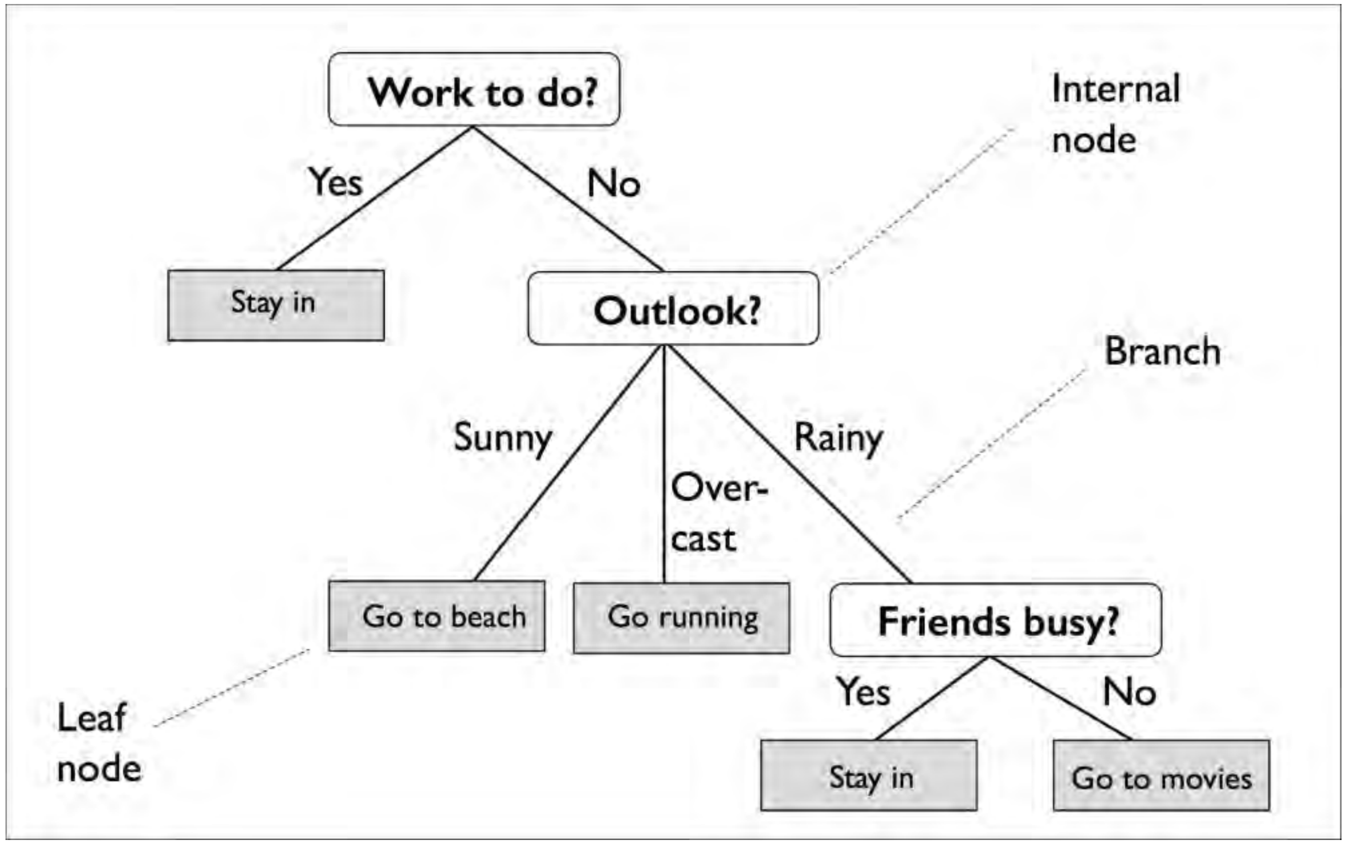
Normally need a **“good” specimen** to perform at their best

Can be available on **paper**, or on a **digital** support

Are **portable** in the field

Can be use **on- and off-line**

The identification process leads to **one name**



Identification

Once upon A time

Identification

The first approaches to digital identification date back to the beginning of the computer era.

The idea was to use the computational opportunities granted by the computers to produce stand-alone, paper printed keys based on databases of characters, getting rid of the constraints of biological classification.

To achieve this goal, DELTA (Description Language for Taxonomy), a structured language to convert scientific descriptions into computer-readable data, was developed.

DELTA was developed since 1971 in Australia, at the Department of entomology of CSIRO (Commonwealth Scientific and Industrial Research Organisation).

Identification

The evolution of internet, both in terms of accessibility and efficiency in the transmission of data, permitted to "mobilise" the identification keys and publish them online.

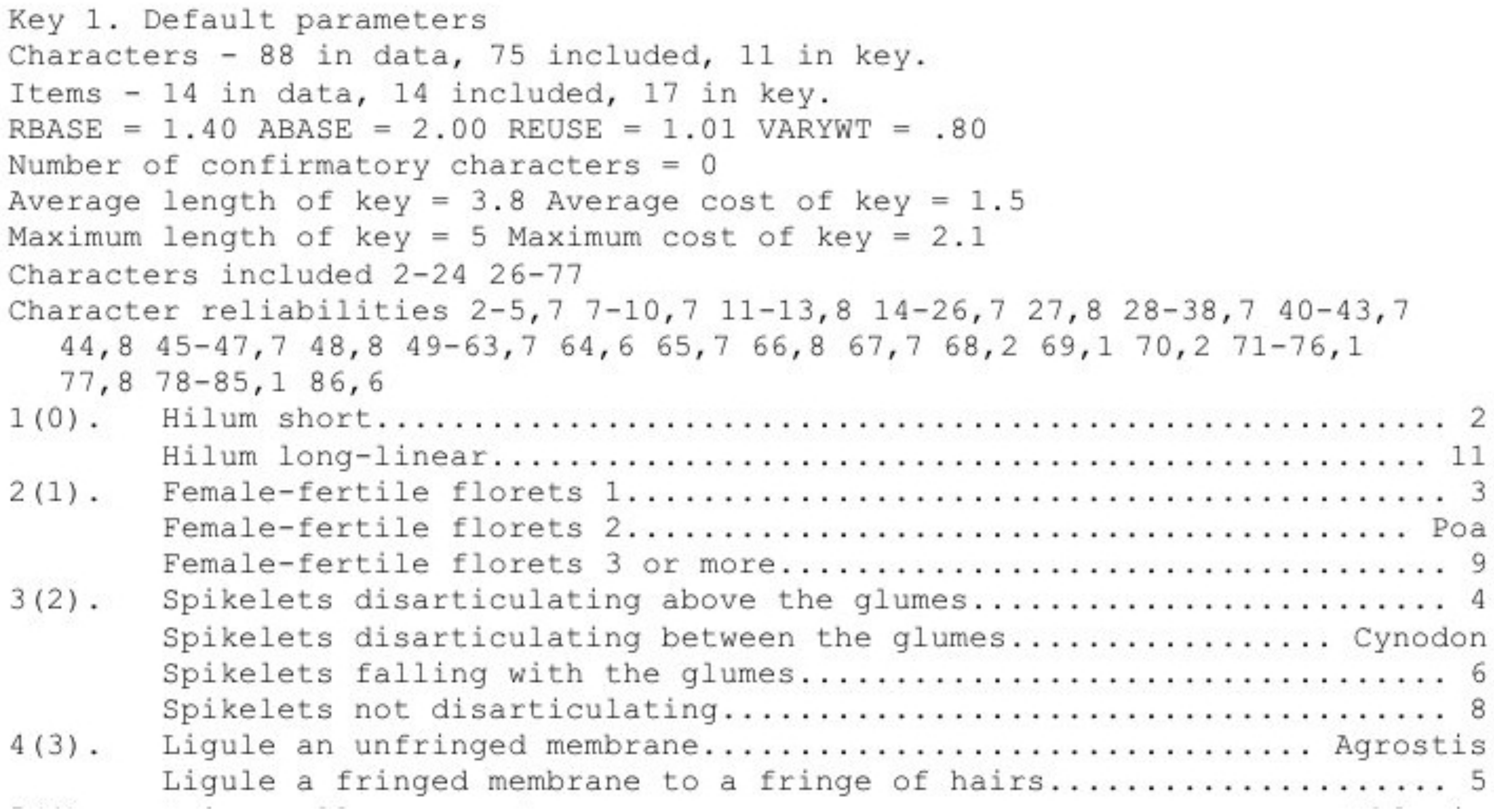
The first approaches to online publication, however, were something like going back to the past. In fact, paper printed keys were simply converted into HTML pages, and published online. This approach is relatively simple, and requires limited skills in the use of HTML tags.

More "evolute" approaches were..... too “evolute” for their times. Hence, the first keys produced by the DELTA suite were simply shared as DELTA files, readable only by someone with the skills to use them, in an another instance of the DELTA suite.

Identification

The DELTA suite is made of several software tools. Program KEY is used to produce digital textual keys.

KEY uses the data stored in DELTA format, selecting them following the rules of a complex algorithm, which mixes author's input and a score based ont he character's relative resolution.



```
Key 1. Default parameters
Characters - 88 in data, 75 included, 11 in key.
Items - 14 in data, 14 included, 17 in key.
RBASE = 1.40 ABASE = 2.00 REUSE = 1.01 VARYWT = .80
Number of confirmatory characters = 0
Average length of key = 3.8 Average cost of key = 1.5
Maximum length of key = 5 Maximum cost of key = 2.1
Characters included 2-24 26-77
Character reliabilities 2-5,7 7-10,7 11-13,8 14-26,7 27,8 28-38,7 40-43,7
   44,8 45-47,7 48,8 49-63,7 64,6 65,7 66,8 67,7 68,2 69,1 70,2 71-76,1
   77,8 78-85,1 86,6
1(0).   Hilum short........................................................ 2
        Hilum long-linear................................................. 11
2(1).   Female-fertile florets 1........................................... 3
        Female-fertile florets 2......................................... Poa
        Female-fertile florets 3 or more................................... 9
3(2).   Spikelets disarticulating above the glumes......................... 4
        Spikelets disarticulating between the glumes................. Cynodon
        Spikelets falling with the glumes.................................. 6
        Spikelets not disarticulating...................................... 8
4(3).   Ligule an unfringed membrane................................ Agrostis
        Ligule a fringed membrane to a fringe of hairs..................... 5
```

Identification

In DELTA, taxa are described by states selected from lists of characters. Characters can have none or two or more states.

5 types of characters are defined:

- ordered multistate (states are arranged following a logical sequence)
- unordered multistate (states are not arranged in a logical sequence)
- integer (1, 2, etc.)
- non-integer (0.2, 1.5, etc.)
- textual (no fixed values)

Identification

DELTA's grammar states that before each character there must be a "#" sign, followed by the number of the character and by a dot. The name of the character follows, optionally with a descriptions written between <>. All is followed by the end-of-line symbol, the slash.

as an example:

#1. Fur <fur is present or not>/

This line is followed by as many lines as the states of the character.

Each state is written starting by its number followed by a dot, the state and its description, and each line is closed by a slash.

Identification

as an example:

1. present <animal with a fur>/
2. absent <animal without a fur>/

Numeric characters have a different grammar. They are stateless, and are followed by the mesure unit, if available.

As an example:

#2. Antennae <length>/ mm/

Textual characters are stateless, and often nameless as well:

#3. <notes>/

Identification

Taxa are described by one or more "instances", each one describes a "variation" of characters inside a taxon.

As an example, a taxon can have organisms with and without fur, so that it will be described by two instances, or records.

These are written following a "#" sign, and start with taxon name, followed by a comment (often used for authors' names):

An example:

Artemisia alba <Turra>/

Other records or instances of the same taxon are written in the same way, but with a plus sign immediately after the "#".

For each character, then, states are listed.

Identification

A state of a taxon is written as character number followed by state number, in the format:

C,S

Different states are separated by a space.

S is the number of the state in the case of multistate characters, or a number in the case of numerical characters.

Three pseudo-values are accepted:

V (variable)

U (undefined)

- (not usable)

Identification

Logical separators are accepted:

/ or

& and

- from to (range)

an example:

species <author>/ 1,1/2<rare> 2,- 3,4-5 4,V

In this case, we have a taxon with the first character which could have two states, but the second is rarer. The second charater is not usable, the thirs ranges from 4 to 5, and the fourth is variable.

Identification

Numerical characters add some issue.

A ranging numerica character is normally written as follows:

C,x-y

However, ranges are not that precise, in nature. Normally there exist a "strict" range, which is the "normality" for the taxon, and a "wide" range, which takes into account a certain deviation from "normality" whcih is rarely observed. This can be expressed as:

C,(1-)2-3(-5)

C,(3-)4(-5)

C,(1-)2-3

Identification

Species species <Auct.>/

1,V 2,3 3,– 4,120

#1. Fur <presence>/

1. present/
2. absent/

#2. Fur's colour/

1. Red/
2. Black/
3. Brown/

#3. Number of fingers in the front legs/

#4. Body's length/ cm/

Identification

Linnaeus (ETI Bioinformatics, Amsterdam, The Nederlands)

It is made of three parts

builder: permits to create and edit an information system on a group of organisms. This information system is made of four modules:

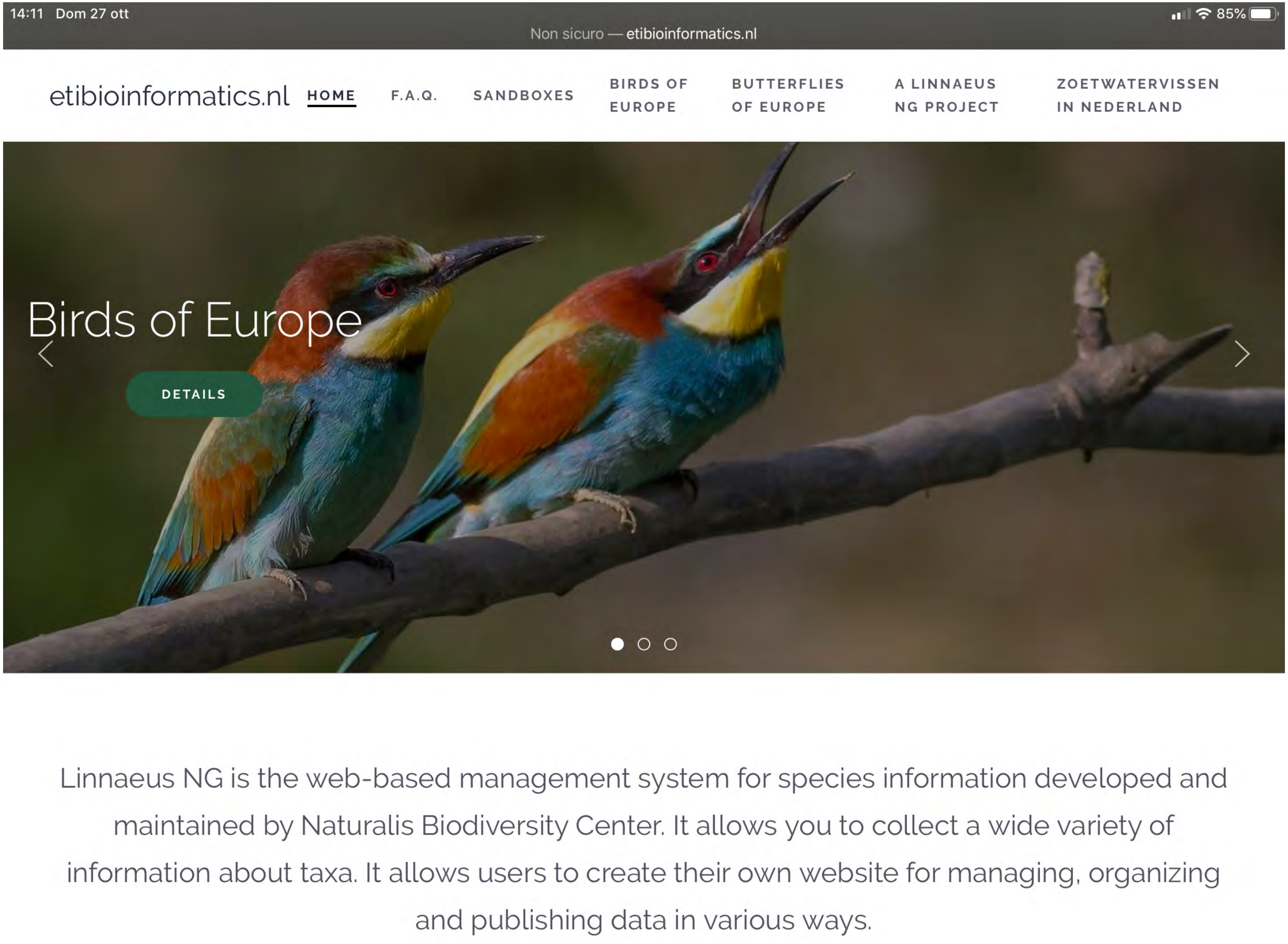
a) taxonomic databases
b) supporting database, to store non-species related information
c) identification tools: visual key, classic key, and multi-access key
d) bio-geographic information system

runtime: permits to store the information system on CD- or DVD-ROMs

publisher: permits to publish the information system on the web

Linneus is now available as a web tool, at the address http://www.etibioinformatics.nl

Identification



Linnaeus NG is the web-based management system for species information developed and maintained by Naturalis Biodiversity Center. It allows you to collect a wide variety of information about taxa. It allows users to create their own website for managing, organizing and publishing data in various ways.

Identification

FRIDA (FRiendly IDentificAtion) was developed since 2003 at the Department of Life Science of the University of Trieste (Italy), to be adopted in projects for the production of multi-authored digital identification keys. Its main aim was to make different experts work together to the same dataset without conflictin with each other.

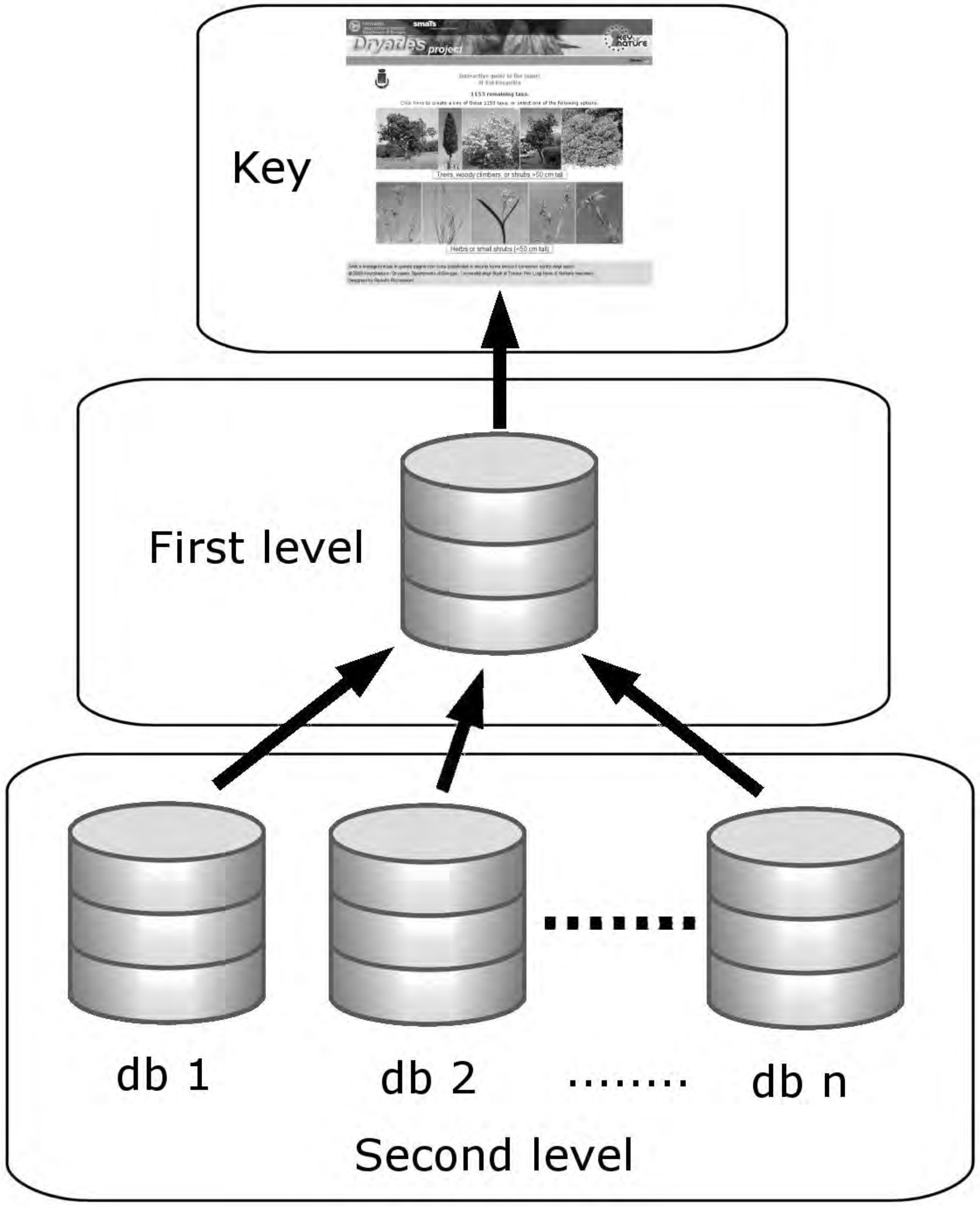
The keys have a dichotomous structure which follows the sequence of identification characters decided by the author(s).

The keys are available in the web, and can be stored on CD- or DVD-Roms, PDAs and Smartphones.

Identification

The database of characters is organised on two levels of information, and both data input and management are made through simple HTML interfaces.

FRIDA can produce keys to any selection of taxa from the database, generating true multi-authored identification keys.



Ma torniamo un attimo indietro…

……e precisamente al 1455.

·Geneſis·

Incipit lib̃ breſith quē nos geneſī dicimuˢ

In principio creauit deus celū cap̄ pᵐ et terram. Terra autem erat inanis et vacua: ⁊ tenebre erant ſup faciē abiſſi: et ſpiritus dei ferebatur ſuper aquas. Dixitq; deus. Fiat lux. Et facta ē lux. Et vidit deus lucem ꝙ eſſet bona: et diuiſit lucem a tenebris· appellauitq; lucem diem et tenebras noctem. Factū q; eſt veſpere ⁊ mane dies vnus. Dixit quoq; deus. Fiat firmamentū in medio aquarū: et diuidat aquas ab aquis. Et fecit deus firmamentū: diuiſitq; aquas que erant ſub firmamento ab hiis que erant ſuper firmamentum: ⁊ factum eſt ita. Vocauitq; deus firmamentū celū: ⁊ factum eſt veſpere et mane dies ſecundus. Dixit vero deus. Congregentur aque que ſub celo ſunt in locum vnū· et appareat arida. Et factum eſt ita. Et vocauit deus aridam terram: cōgregationeſq; aquarum appellauit maria. Et vidit deus ꝙ eſſet bonū· et ait. Germinet terra herbā virentem et facientem ſemen: et lignū pomiferū faciens fructum iuxta genus ſuū: cuius ſemen in ſemetipſo ſit ſuper terram. Et factum eſt ita. Et protulit terra herbam virentem et afferentē ſemen iuxta genus ſuū: lignūq; faciens fructū et habēs vnūquodq; ſementē ſcdm ſpeciē ſuā. Et vidit deus ꝙ eſſet bonū: et factū ē veſpere et mane dies tercius. Dixitq; aūt deus. Fiant luminaria in firmamēto celi· ⁊ diuidāt diem ac noctē: ⁊ ſint in ſigna ⁊ tēpora· ⁊ dies ⁊ annos: ut luceāt in firmamēto celi et illuminēt terrā. Et factū eſt ita. Fecitq; deus duo luminaria magna: luminare maius ut p̄eſſet diei et luminare min⁹ ut p̄eſſet nocti: ⁊ ſtellas· ⁊ poſuit eas in firmamēto celi ut lucerent ſup terrā: et p̄eſſent diei ac nocti: ⁊ diuiderēt lucem ac tenebras. Et vidit de⁹ ꝙ eſſet bonū: et factū ē veſpere et mane dies quart⁹. Dixit etiam deus. Producant aque reptile anime viuentis et volatile ſup terram: ſub firmamēto celi. Creauitq; deus cete grandia· et omnē animā viuentem atq; motabilem quā produxerant aque in ſpecies ſuas: ⁊ omne volatile ſecundū genus ſuū. Et vidit deus ꝙ eſſet bonū: benedixitq; ei dicens. Creſcite et multiplicamini· et replete aquas maris: aueſq; multiplicentur ſuper terram. Et factū ē veſpere ⁊ mane dies quītus. Dixit quoq; deus. Producat terra animā viuentem in genere ſuo: iumenta ⁊ reptilia· ⁊ beſtias terre· ſecundū ſpecies ſuas. Factū ē ita. Et fecit deus beſtias terre iuxta ſpecies ſuas: iumenta ⁊ omne reptile terre in genere ſuo. Et vidit deus ꝙ eſſet bonū: et ait. Faciam⁹ hominē ad ymaginē ⁊ ſilitudinē noſtrā· ⁊ p̄ſit piſcib; maris· ⁊ volatilib; celi· ⁊ beſtiis· vniūſeq; terre: omīq; reptili qđ mouet̄ ī terra. Et creauit deus hominē ad ymaginē et ſimilitudinē ſuam: ad ymaginem dei creauit illū: maſculū et feminā creauit eos. Benedixitq; illis deus· et ait. Creſcite et multiplicamini ⁊ replete terram· et ſubicite eam: ⁊ dominamini piſcibus maris· ⁊ volatilibus celi: ⁊ vniuerſis animātibus que mouentur ſup terrā. Dixitq; deus. Ecce dedi vobis omnē herbam afferentem ſemen ſup terram· et vniūſa ligna que habēt ī ſemetipſis ſemētē generis ſui: ut ſint vobis ī eſcā· ⁊ cūctis aīantibus terre· omīq; volucri celi ⁊ vniuerſis q̄ mouētur in terra· et ī quibus ē anima viuēs: ut habeāt ad veſcendū. Et factū eſt ita. Viditq; deus cuncta que fecerat: et erāt valde bona.

Identification

Digital identification keys can be produced by following any sequence of characters, not only that imposed by biologic classification. A computer can in fact shuffle the characters, on the basis of users' requests.

As an example, it is possible to produce keys which prioritize colours (of leaves, flowers and fruits) or shapes, hence creating different keys to the same group of organisms. These keys can have different applications for different target users and in different time frame.

Furthermore, it is possible to introduce in the keys features which are normally absent in "classic" keys, such as ecological, and distributional informations, hence contextualising the lists of organisms.

Identification

The publication of identification keys in paper printed journals and books forced authors to follow the logical scheme of biological classification, especially while dealing with big taxa.

Human brain, even when well trained, can not organise in a data table all the features which are necessary to identify all the organisms e.g. in a country, and to organise these data as a computer, when well programmed, can.

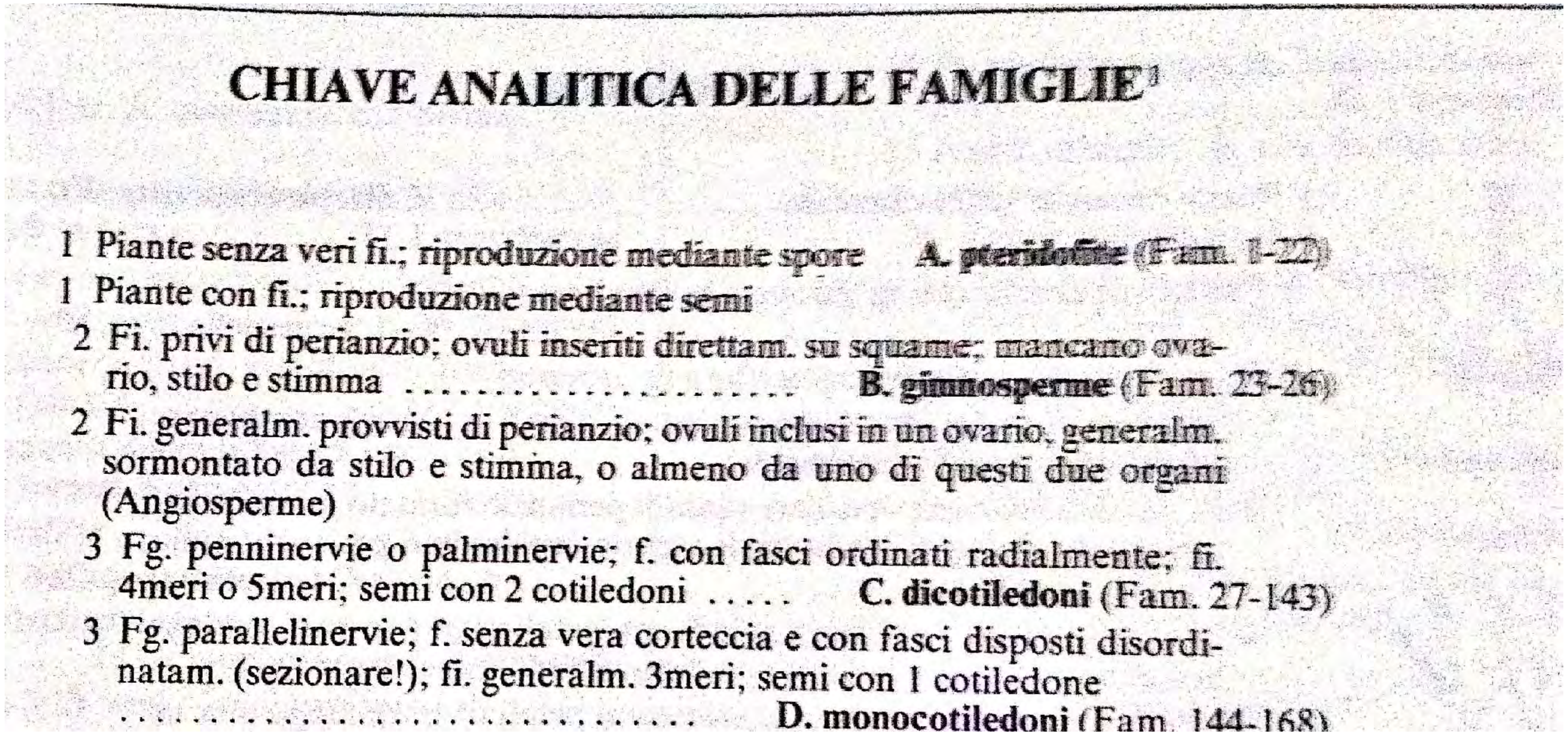
Hence, it is necessary, for a scientist, to split large groups into smaller ones, until reaching "manageable" sizes, normally at genus level. To do this, the better way is to follow biologic classification.

"Classic" paper printed keys first lead to orders, then to families and genera. At the end, for each genus, a key to the species is provided.

Identification

In these keys, biological classification is the key to access species' names. Hence, the misunderstanding of considering identification and classification as synonyms.

Features which are used to identify the higher taxa are, however, normally difficult. Hence, "classic" keys are intrinsically difficult.



CHIAVE ANALITICA DELLE FAMIGLIE[1]

1 Piante senza veri fi.; riproduzione mediante spore **A. pteridofite** (Fam. 1-22)
1 Piante con fi.; riproduzione mediante semi
 2 Fi. privi di perianzio; ovuli inseriti direttam. su squame; mancano ovario, stilo e stimma .... **B. gimnosperme** (Fam. 23-26)
 2 Fi. generalm. provvisti di perianzio; ovuli inclusi in un ovario, generalm. sormontato da stilo e stimma, o almeno da uno di questi due organi (Angiosperme)
  3 Fg. penninervie o palminervie; f. con fasci ordinati radialmente; fi. 4meri o 5meri; semi con 2 cotiledoni ..... **C. dicotiledoni** (Fam. 27-143)
  3 Fg. parallelinervie; f. senza vera corteccia e con fasci disposti disordinatam. (sezionare!); fi. generalm. 3meri; semi con 1 cotiledone .... **D. monocotiledoni** (Fam. 144-168)

Identification

Ovary inferior	*Order*	*Ovary superior*
Carpels 4, free	*Family*	*Carpels 5, fused*
Stamens hairy	*Genus*	*Stamens glabrous*
Fruit dehiscent	*Species*	*Fruit non dehiscent*

Identification

Flowers red, leaves divided

~~Order~~

~~Family~~

~~Genus~~

~~Species~~

Flowers white, leaves entire

Identification

Classic keys are printed on paper, and can not be modified without a new printed edition.

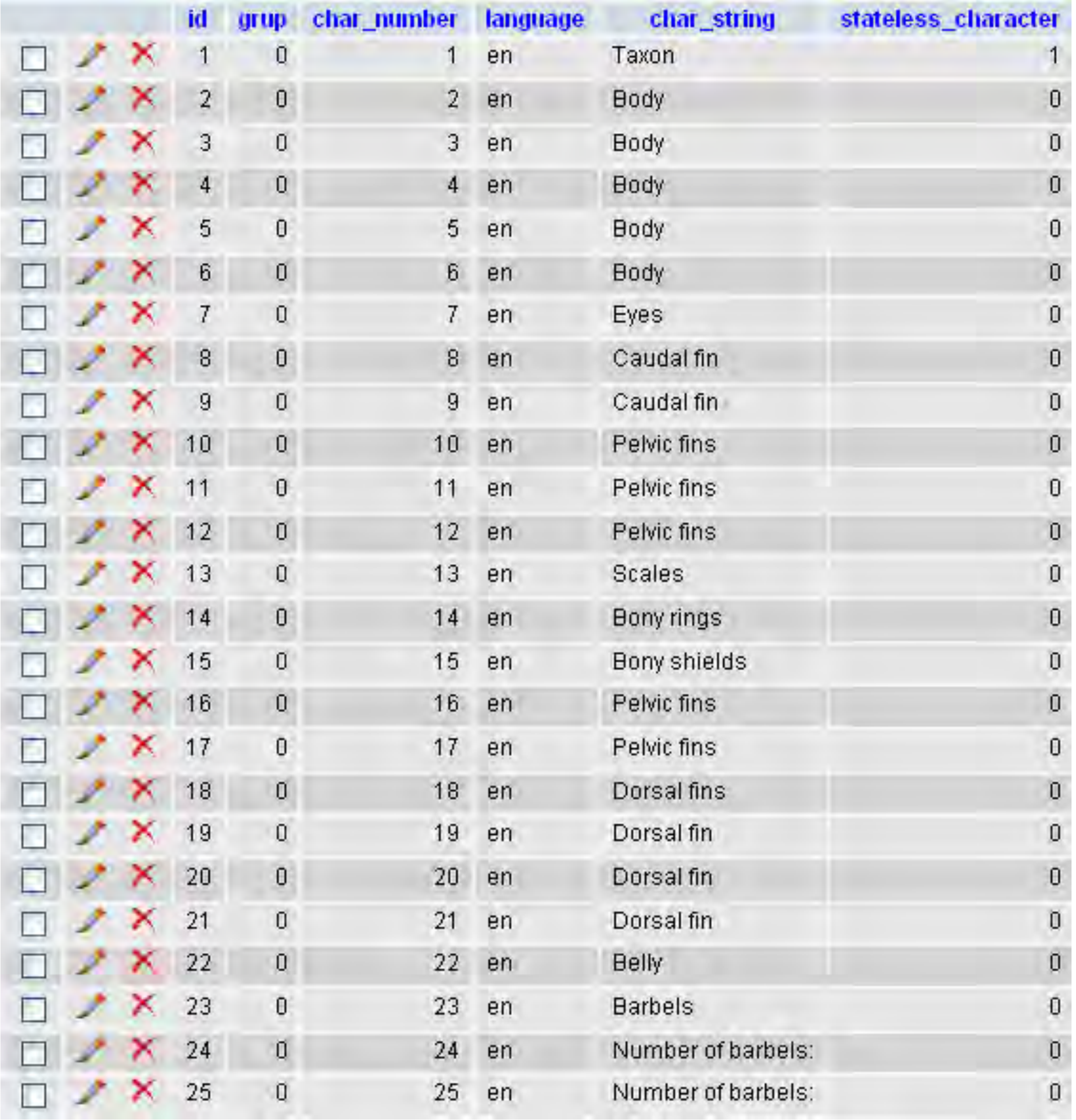
Nomenclatural changes, progress in exploration, discovery of new species, often render a classic key outdated within a few years.

Computerised systems can be updated and corrected in real time.

Identification

			TP	C001	C002	C003	C004	C005	C006	C007	C008	C009	C010	C011	C012	C013	C014	C015	C016	C017	C018	C019
☐	✎	✗	1	61	Platichthys flesus (Linnaeus, 1758)	2	1	1	1	2	2	1	2	1	2	1	1	2	2	1	2	1
☐	✎	✗	2	29	Gobio benacensis (Pollini, 1816)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	30	Tinca tinca (Linnaeus, 1758)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	31	Barbus plebejus Bonaparte, 1839	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	32	Barbus caninus Bonaparte, 1839	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	33	Cyprinus carpio Linnaeus, 1758	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	35	Cyprinus carpio Linnaeus, 1758	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	36	Carassius carassius (Linnaeus, 1758)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	37	Carassius auratus (Linnaeus, 1758)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	38	Chondrostoma soetta Bonaparte, 1840	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	39	Chondrostoma nasus (Linnaeus, 1758)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	40	Chondrostoma genei (Bonaparte, 1839)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	41	Phoxinus phoxinus (Linnaeus, 1758)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1
☐	✎	✗	2	42	Ctenopharyngodon idellus (Valenciennes, 1844)	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	0	1

Identification



			id	grup	char_number	language	char_string	stateless_character
☐	✎	✗	1	0	1	en	Taxon	1
☐	✎	✗	2	0	2	en	Body	0
☐	✎	✗	3	0	3	en	Body	0
☐	✎	✗	4	0	4	en	Body	0
☐	✎	✗	5	0	5	en	Body	0
☐	✎	✗	6	0	6	en	Body	0
☐	✎	✗	7	0	7	en	Eyes	0
☐	✎	✗	8	0	8	en	Caudal fin	0
☐	✎	✗	9	0	9	en	Caudal fin	0
☐	✎	✗	10	0	10	en	Pelvic fins	0
☐	✎	✗	11	0	11	en	Pelvic fins	0
☐	✎	✗	12	0	12	en	Pelvic fins	0
☐	✎	✗	13	0	13	en	Scales	0
☐	✎	✗	14	0	14	en	Bony rings	0
☐	✎	✗	15	0	15	en	Bony shields	0
☐	✎	✗	16	0	16	en	Pelvic fins	0
☐	✎	✗	17	0	17	en	Pelvic fins	0
☐	✎	✗	18	0	18	en	Dorsal fins	0
☐	✎	✗	19	0	19	en	Dorsal fin	0
☐	✎	✗	20	0	20	en	Dorsal fin	0
☐	✎	✗	21	0	21	en	Dorsal fin	0
☐	✎	✗	22	0	22	en	Belly	0
☐	✎	✗	23	0	23	en	Barbels	0
☐	✎	✗	24	0	24	en	Number of barbels:	0
☐	✎	✗	25	0	25	en	Number of barbels:	0

			id	grup	char_number	state_number	language	state_string	image_id
☐	✎	✗	1	0	2	2	en	clearly flat	NULL
☐	✎	✗	2	0	2	1	en	not clearly flat	NULL
☐	✎	✗	3	0	3	1	en	not polygonal in section	NULL
☐	✎	✗	4	0	3	2	en	polygonal in section	NULL
☐	✎	✗	5	0	4	2	en	very long	NULL
☐	✎	✗	6	0	4	1	en	not very long	NULL
☐	✎	✗	7	0	5	2	en	much longer than high	NULL
☐	✎	✗	8	0	5	1	en	not much longer than high	NULL
☐	✎	✗	9	0	6	2	en	not like a snake	NULL
☐	✎	✗	10	0	6	1	en	like a snake	NULL
☐	✎	✗	11	0	7	2	en	both placed on the pigmented side	NULL
☐	✎	✗	12	0	7	1	en	placed on both sides	NULL
☐	✎	✗	13	0	8	1	en	not heterocercal	NULL
☐	✎	✗	14	0	8	2	en	heterocercal (with asymmetric lobes)	NULL
☐	✎	✗	15	0	9	2	en	with rounded margin	NULL
☐	✎	✗	16	0	9	1	en	with symmetric lobes or with straight margin	NULL
☐	✎	✗	17	0	10	2	en	absent	NULL
☐	✎	✗	18	0	10	1	en	present, sometimes reduced at spines	NULL
☐	✎	✗	19	0	11	2	en	not reduced at spines	NULL
☐	✎	✗	20	0	11	1	en	reduced at spines	NULL
☐	✎	✗	21	0	12	2	en	united to form a sucker	NULL
☐	✎	✗	22	0	12	1	en	separated	NULL
☐	✎	✗	23	0	13	2	en	absent	NULL
☐	✎	✗	24	0	13	1	en	present	NULL
☐	✎	✗	25	0	14	2	en	absent	NULL

Identification

The larger a group of organisms is, the higher the number of characters required to identify them, the more difficult for an user to identify them.

Computerised tools permit to reduce the set of organisms using different combinations of morphological, ecological, and distributional characters.

Identification

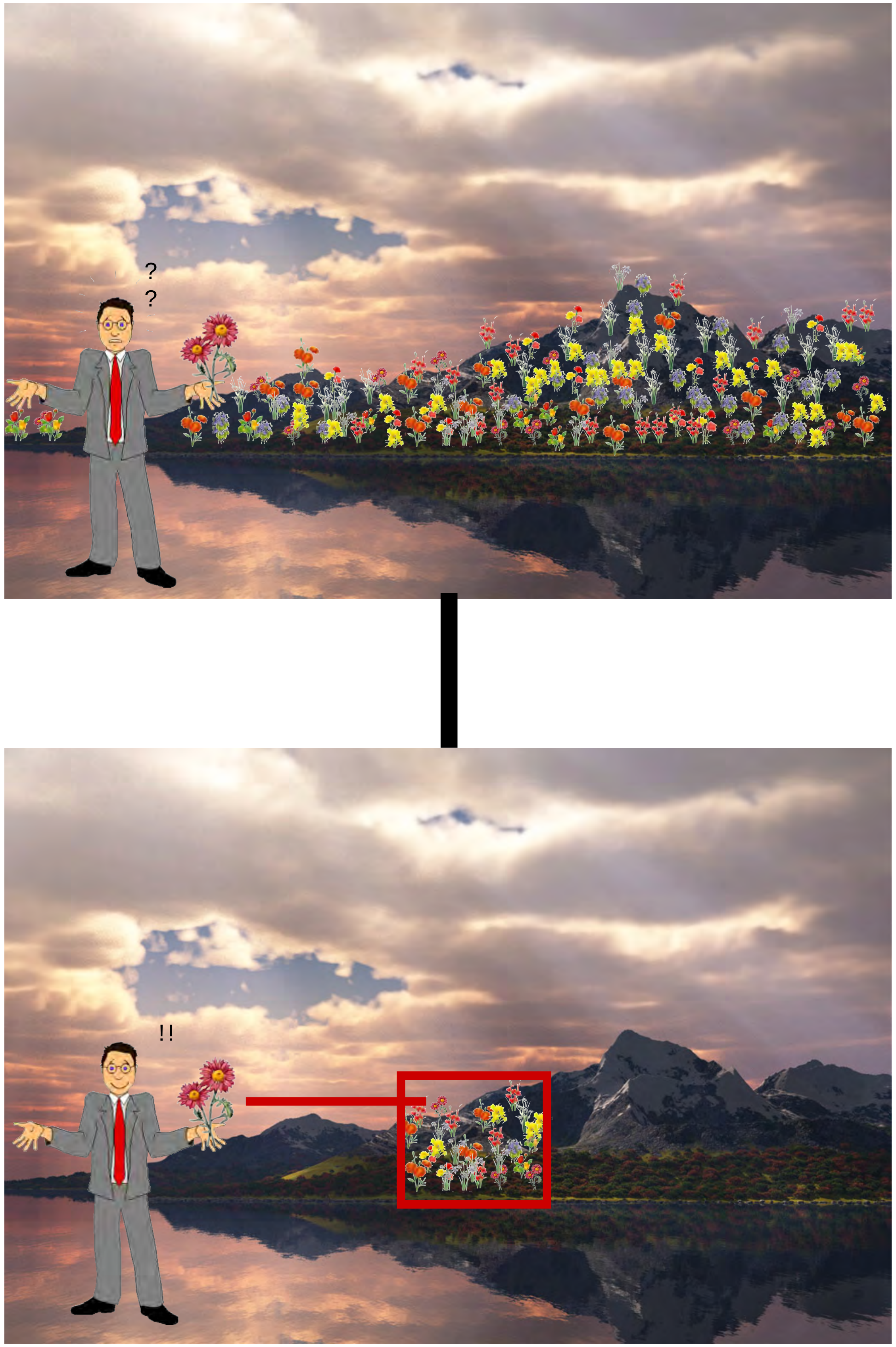
Spatial contextualisation.

National, regional, or local checklists normally contain an high number of congeneric organisms, which are often very similar.

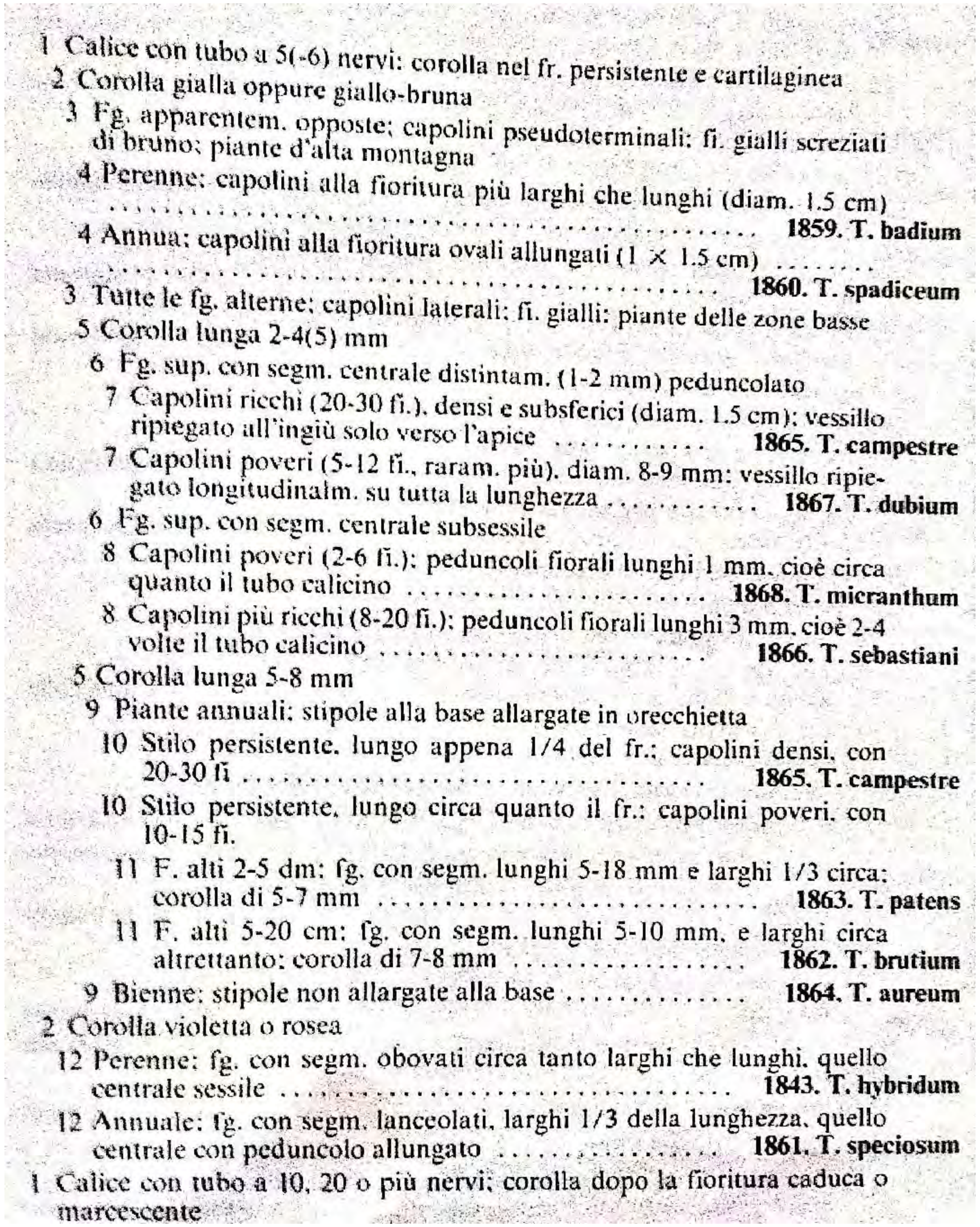
Their identification is tuhs difficult, since it requires the observation of "complex" features.

As an example, the genus *Trifolium* in Italy counts ca. 100 infrageneric taxa.

Reducing the key to a subset of the whole flora makes things far easier.



Identification



1 Calice con tubo a 5(-6) nervi: corolla nel fr. persistente e cartilaginea
  2 Corolla gialla oppure giallo-bruna
    3 Fg. apparentem. opposte; capolini pseudoterminali; fi. gialli screziati di bruno; piante d'alta montagna
      4 Perenne: capolini alla fioritura più larghi che lunghi (diam. 1.5 cm) .... **1859. T. badium**
      4 Annua: capolini alla fioritura ovali allungati (1 × 1.5 cm) .... **1860. T. spadiceum**
    3 Tutte le fg. alterne; capolini laterali: fi. gialli: piante delle zone basse
      5 Corolla lunga 2-4(5) mm
        6 Fg. sup. con segm. centrale distintam. (1-2 mm) peduncolato
          7 Capolini ricchi (20-30 fi.), densi e subsferici (diam. 1.5 cm); vessillo ripiegato all'ingiù solo verso l'apice .... **1865. T. campestre**
          7 Capolini poveri (5-12 fi., raram. più), diam. 8-9 mm; vessillo ripiegato longitudinalm. su tutta la lunghezza .... **1867. T. dubium**
        6 Fg. sup. con segm. centrale subsessile
          8 Capolini poveri (2-6 fi.); peduncoli fiorali lunghi 1 mm, cioè circa quanto il tubo calicino .... **1868. T. micranthum**
          8 Capolini più ricchi (8-20 fi.); peduncoli fiorali lunghi 3 mm, cioè 2-4 volte il tubo calicino .... **1866. T. sebastiani**
      5 Corolla lunga 5-8 mm
        9 Piante annuali: stipole alla base allargate in orecchietta
          10 Stilo persistente, lungo appena 1/4 del fr.; capolini densi, con 20-30 fi .... **1865. T. campestre**
          10 Stilo persistente, lungo circa quanto il fr.; capolini poveri, con 10-15 fi.
            11 F. alti 2-5 dm: fg. con segm. lunghi 5-18 mm e larghi 1/3 circa; corolla di 5-7 mm .... **1863. T. patens**
            11 F. alti 5-20 cm: fg. con segm. lunghi 5-10 mm, e larghi circa altrettanto; corolla di 7-8 mm .... **1862. T. brutium**
        9 Bienne: stipole non allargate alla base .... **1864. T. aureum**
  2 Corolla violetta o rosea
    12 Perenne: fg. con segm. obovati circa tanto larghi che lunghi, quello centrale sessile .... **1843. T. hybridum**
    12 Annuale: fg. con segm. lanceolati, larghi 1/3 della lunghezza, quello centrale con peduncolo allungato .... **1861. T. speciosum**
1 Calice con tubo a 10, 20 o più nervi; corolla dopo la fioritura caduca o marcescente

Identification

Normally, congeneric taxa rarely rarely occur in great number in the same, area, if it is "small enough".

The size obviously depends on the group of organisms which are investigated.

As an example, in the garden of a school, there will be hardly more than two or three species of the genus *Trifolium*.

Contextualising a digital identification key to an area permits to get rid of all the organisms which are known, because of their distribution and ecological constraints, not to occur in that area.

The contextualisation, as far as digital keys are concerned, can be obtained by following three approaches:

- a priori, by producing small digital keys for different areas.
- a posteriori, by producing keys which are context aware, normally by using the GPS devices which are present on mobile devices
- mixed, by producing different keys for contiguous areas, and asking the user to select the area for which he is performing the identification.

Identification

Traditional keys contain a large amount of information which is frozen into the logical structure build by their author(s).

Computerised floras are “elastic”. They can generate, from a single dataset, a wealth of different products, which would have required a huge amount of work in the past.

Identification

Often a table of characters is identified as an identification key. This "vision" derives from the history of digital identification, which begone with digital keys made of a data table embedded in the software for the identification.

The heritage of that time is a great amount of digital keys which do not "communicate" with each other, so that their data are "lost", no more reusable.

Nowadays, this approach is outdated, but this confusion is still present.

ICT permits now to build big data tables. Ideally, one table could contain a record for each organism in the world.

These data tables can constantly be updated with new records and characters, while their instances could be used to produce an ideally infinite number of digital identification keys.

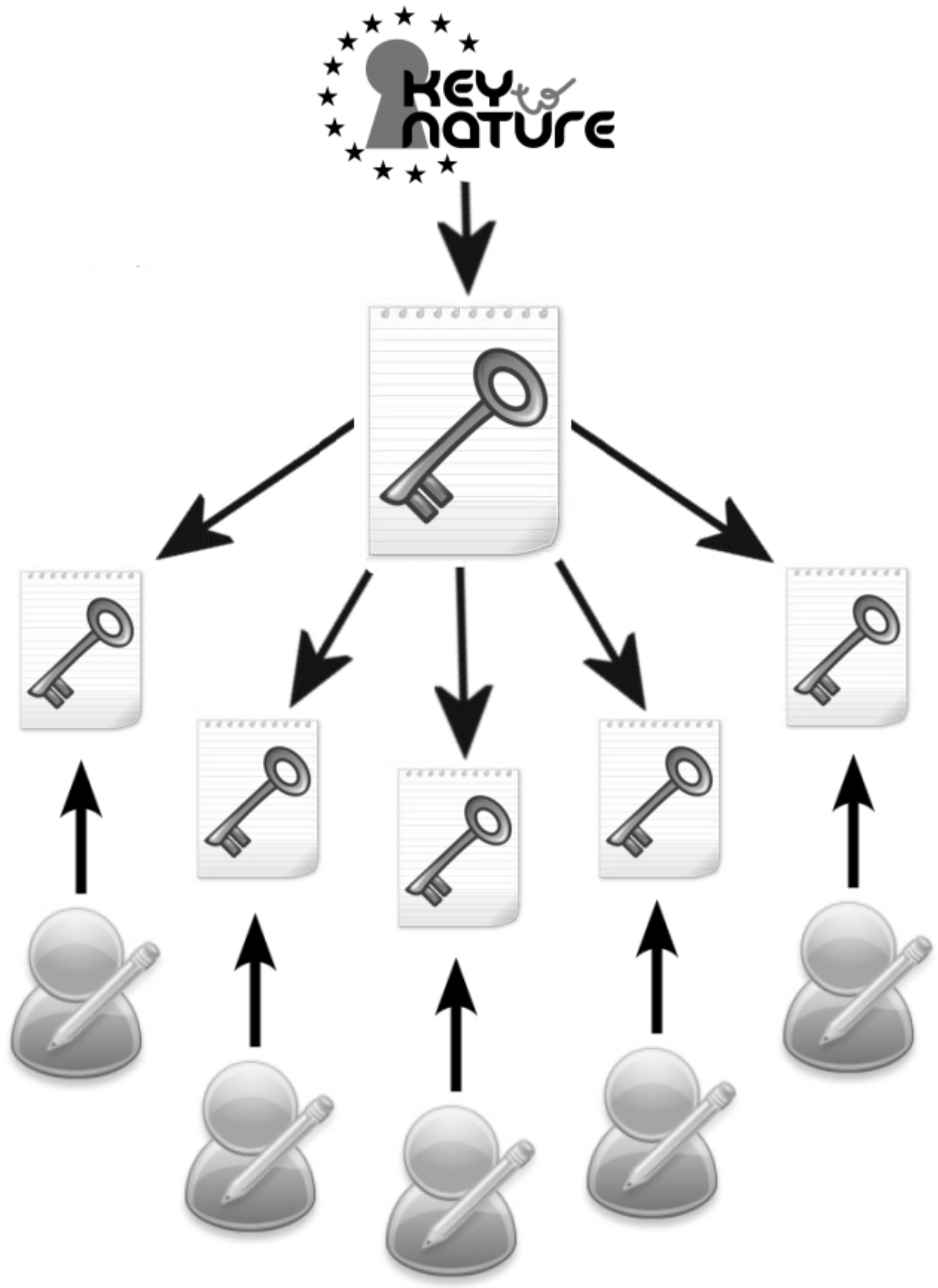
Identification

Digital keys based on databases are “accumulative”.

A small database can be the starting point for future expansions.

Identification

Users can also "personalize" their identification keys with user-generated content, by using ad-hoc softwares, or simply the instruments of the web 2.0.



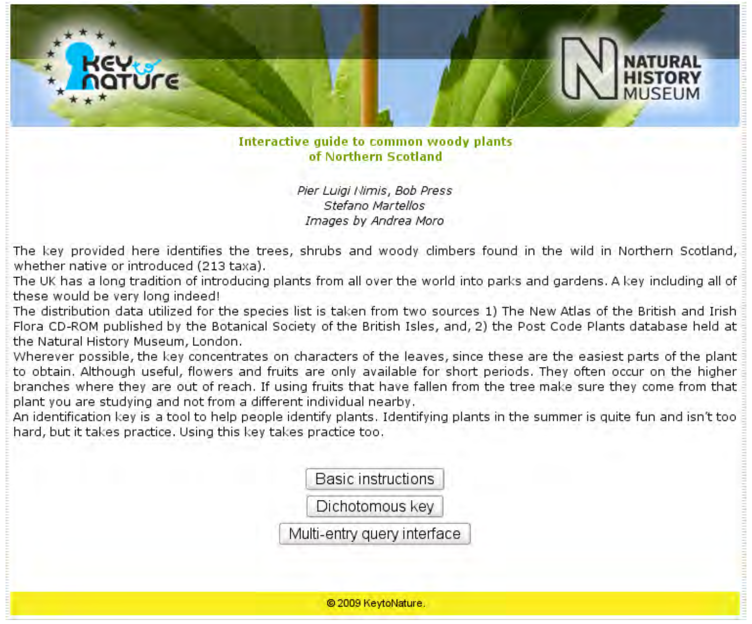
Identification

The result of a computer elaboration can be stored on different digital media, or travel through the internet.

A key can be stored on CD- or DVD-ROMs, as well as on memory cards, or used on-line when an internet connection is available.

They can be also printed on digital books, or on paper as well.

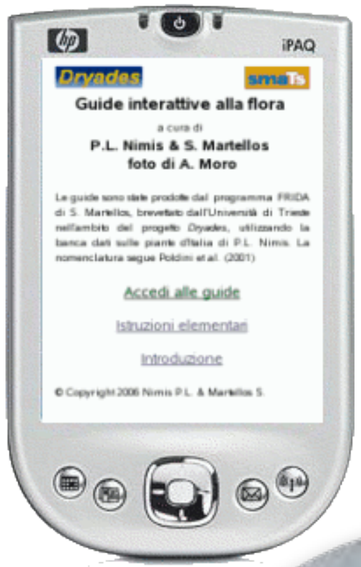
Identification



**Interactive guide to common woody plants
of Northern Scotland**

*Pier Luigi Nimis, Bob Press
Stefano Martellos
Images by Andrea Moro*

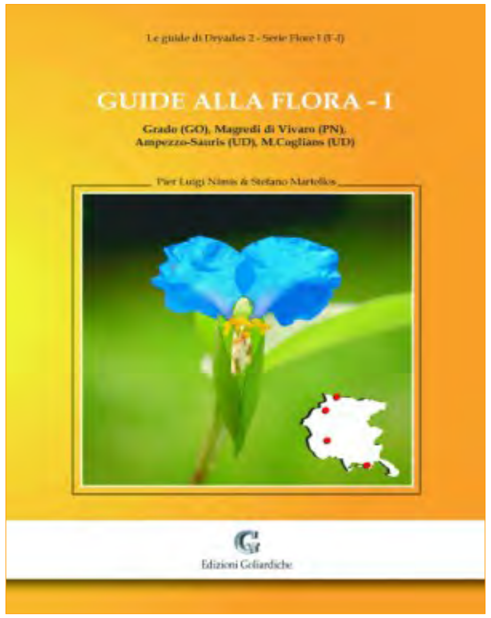
The key provided here identifies the trees, shrubs and woody climbers found in the wild in Northern Scotland, whether native or introduced (213 taxa).
The UK has a long tradition of introducing plants from all over the world into parks and gardens. A key including all of these would be very long indeed!
The distribution data utilized for the species list is taken from two sources 1) The New Atlas of the British and Irish Flora CD-ROM published by the Botanical Society of the British Isles, and, 2) the Post Code Plants database held at the Natural History Museum, London.
Wherever possible, the key concentrates on characters of the leaves, since these are the easiest parts of the plant to obtain. Although useful, flowers and fruits are only available for short periods. They often occur on the higher branches where they are out of reach. If using fruits that have fallen from the tree make sure they come from that plant you are studying and not from a different individual nearby.
An identification key is a tool to help people identify plants. Identifying plants in the summer is quite fun and isn't too hard, but it takes practice. Using this key takes practice too.



Guide interattive alla flora

a cura di
**P.L. Nimis & S. Martellos
foto di A. Moro**

Le guide sono state prodotte dal programma FRIDA di S. Martellos, brevettato dall'Università di Trieste nell'ambito del progetto *Dryades*, utilizzando la banca dati sulle piante d'Italia di P.L. Nimis. La nomenclatura segue Poldini et al. (2001)


Le guide di Dryades 2 - Serie Flore I (F-I)

GUIDE ALLA FLORA - I

Grado (GO), Magredi di Vivaro (PN), Ampezzo-Sauris (UD), M.Coglians (UD)

Pier Luigi Nimis & Stefano Martellos

Edizioni Goliardiche

Structures

fixed pathway keys: identification path was fixed by the author(s)

These keys can be:

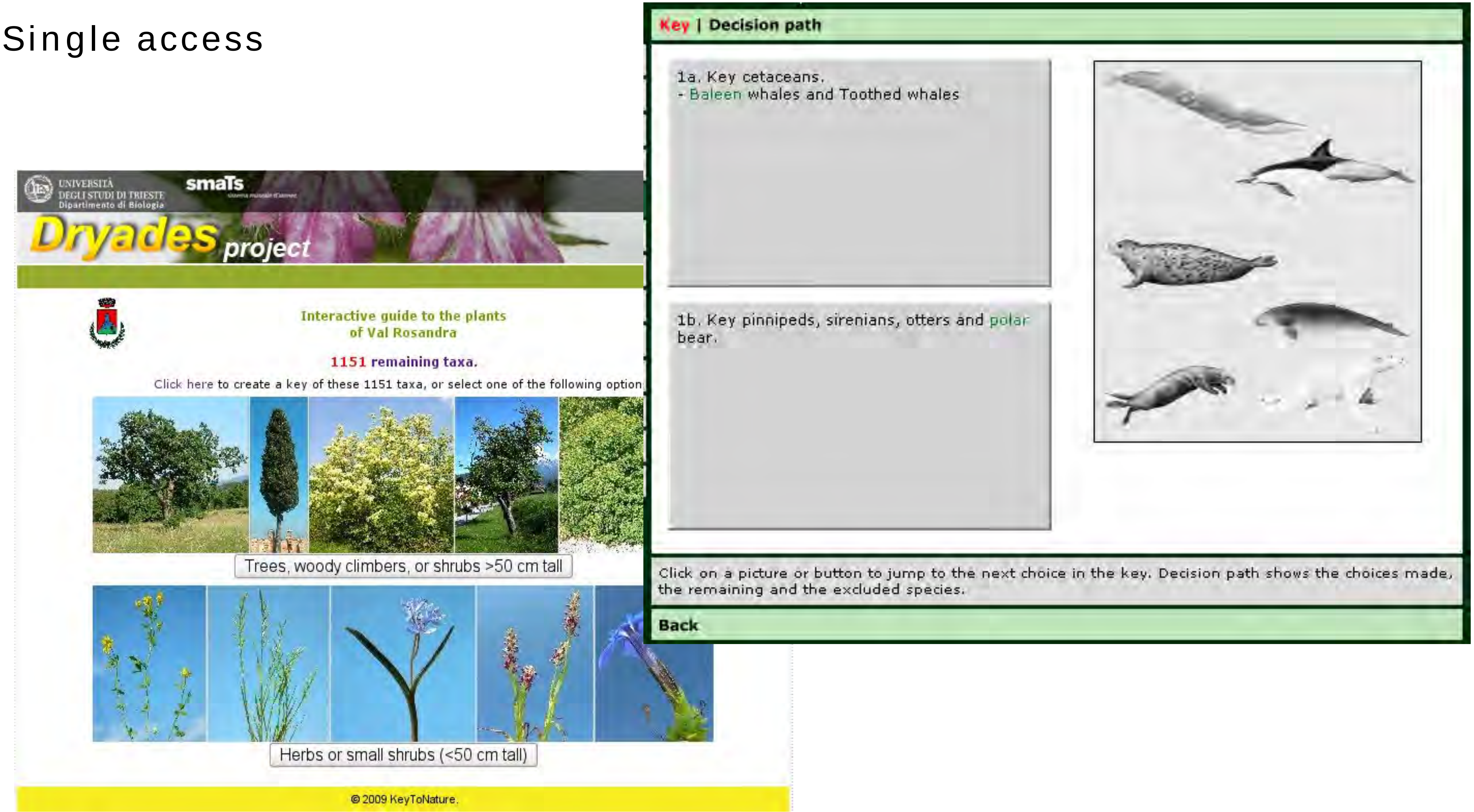
- **static**, if published in the web as simple html pages, or on paper, or when used by single-access interfaces that simply follow the identification path
- **dynamic**, when used through an interface, which, disrupting the original fixed structure, builds a new structure on the basis of user’s input.

free pathway keys: the user is free to choose which character to use in each step of the identification process

These keys are used through multi-access query interfaces, and are always **dynamic**

Interfaces

Single access



Interactive guide to the plants
of Val Rosandra

1151 remaining taxa.

Click here to create a key of these 1151 taxa, or select one of the following option

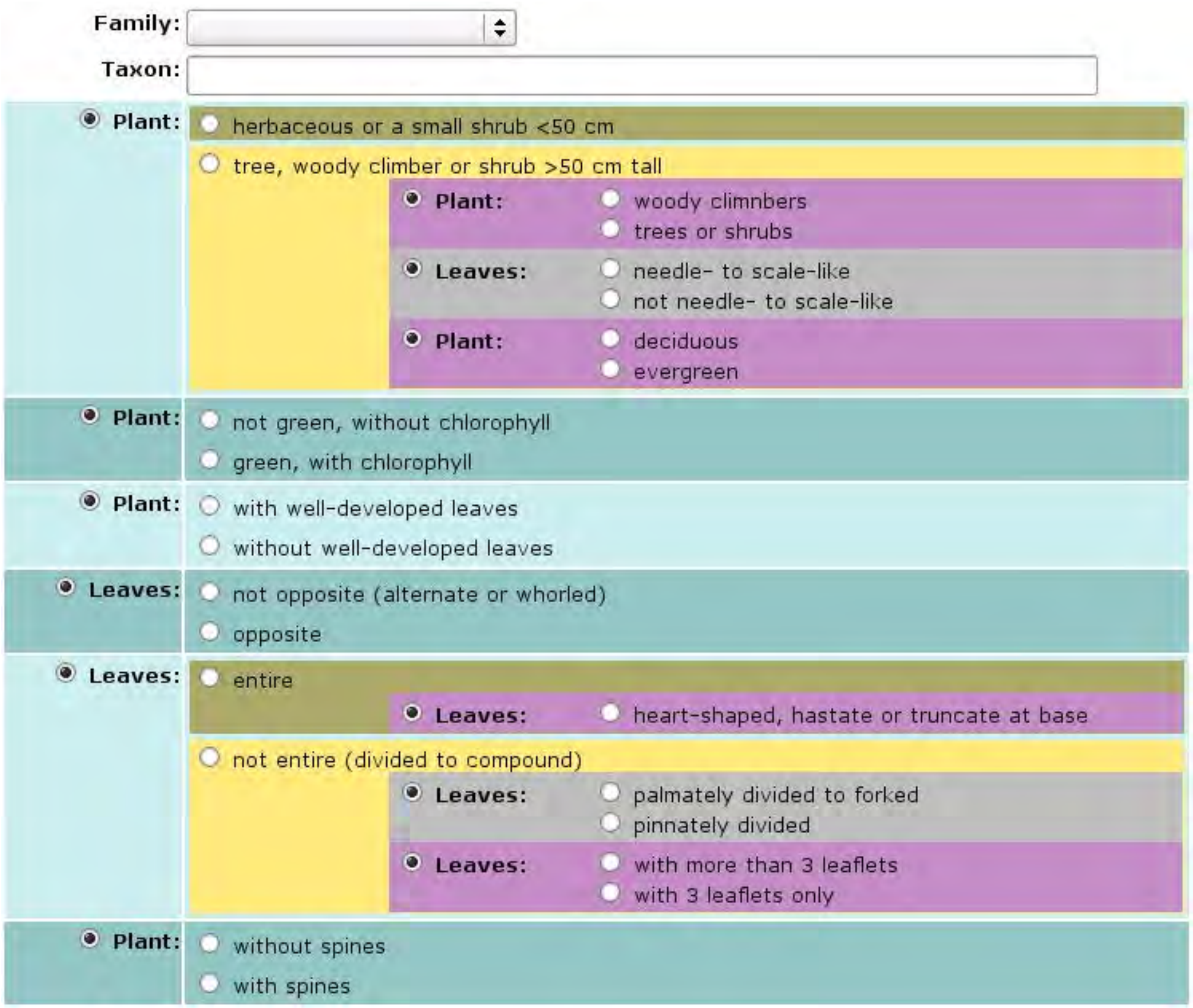
© 2009 KeyToNature.

1a. Key cetaceans.
- Baleen whales and Toothed whales

1b. Key pinnipeds, sirenians, otters and polar bear.

Click on a picture or button to jump to the next choice in the key. Decision path shows the choices made, the remaining and the excluded species.

Multi-entry

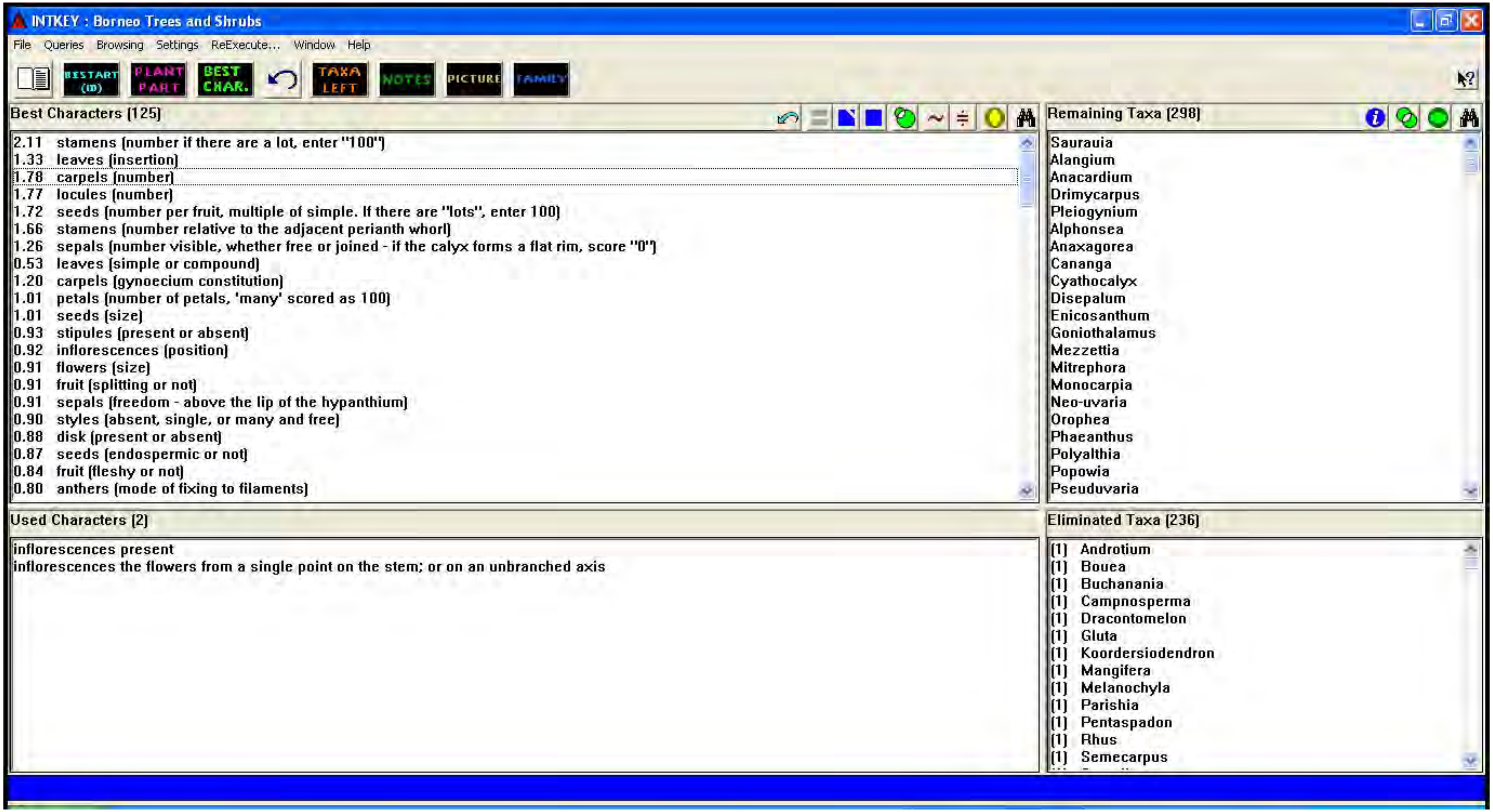


Family:

Taxon:

- ● **Plant:**
  - ○ herbaceous or a small shrub <50 cm
  - ○ tree, woody climber or shrub >50 cm tall
    - ● **Plant:** ○ woody climnbers ○ trees or shrubs
    - ● **Leaves:** ○ needle- to scale-like ○ not needle- to scale-like
    - ● **Plant:** ○ deciduous ○ evergreen
- ● **Plant:**
  - ○ not green, without chlorophyll
  - ○ green, with chlorophyll
- ● **Plant:**
  - ○ with well-developed leaves
  - ○ without well-developed leaves
- ● **Leaves:**
  - ○ not opposite (alternate or whorled)
  - ○ opposite
- ● **Leaves:**
  - ○ entire
    - ● **Leaves:** ○ heart-shaped, hastate or truncate at base
  - ○ not entire (divided to compound)
    - ● **Leaves:** ○ palmately divided to forked ○ pinnately divided
    - ● **Leaves:** ○ with more than 3 leaflets ○ with 3 leaflets only
- ● **Plant:**
  - ○ without spines
  - ○ with spines

Free-access



Best Characters (125)

2.11 stamens (number if there are a lot, enter "100")
1.33 leaves (insertion)
1.78 carpels (number)
1.77 locules (number)
1.72 seeds (number per fruit, multiple of simple. If there are "lots", enter 100)
1.66 stamens (number relative to the adjacent perianth whorl)
1.26 sepals (number visible, whether free or joined - if the calyx forms a flat rim, score "0")
0.53 leaves (simple or compound)
1.20 carpels (gynoecium constitution)
1.01 petals (number of petals, 'many' scored as 100)
1.01 seeds (size)
0.93 stipules (present or absent)
0.92 inflorescences (position)
0.91 flowers (size)
0.91 fruit (splitting or not)
0.91 sepals (freedom - above the lip of the hypanthium)
0.90 styles (absent, single, or many and free)
0.88 disk (present or absent)
0.87 seeds (endospermic or not)
0.84 fruit (fleshy or not)
0.80 anthers (mode of fixing to filaments)

Remaining Taxa (298)

Saurauia
Alangium
Anacardium
Drimycarpus
Pleiogynium
Alphonsea
Anaxagorea
Cananga
Cyathocalyx
Disepalum
Enicosanthum
Goniothalamus
Mezzettia
Mitrephora
Monocarpia
Neo-uvaria
Orophea
Phaeanthus
Polyalthia
Popowia
Pseuduvaria

Used Characters (2)

inflorescences present
inflorescences the flowers from a single point on the stem; or on an unbranched axis

Eliminated Taxa (236)

(1) Androtium
(1) Bouea
(1) Buchanania
(1) Campnosperma
(1) Dracontomelon
(1) Gluta
(1) Koordersiodendron
(1) Mangifera
(1) Melanochyla
(1) Parishia
(1) Pentaspadon
(1) Rhus
(1) Semecarpus

The “best” character dilemma

Some questions:

Which is the “best” character in an identification key?

Which is the “best” sequence of characters in an identification key?

Does it exits ONE “best” sequence for characters in an identification key?

According to Dallawitz, Paine & Zurcher:

“the best 2-state characters are those that divide the remaining taxa into groups that are as nearly equal as possible”.

The number of steps required to reach each organism in the key will be equal to

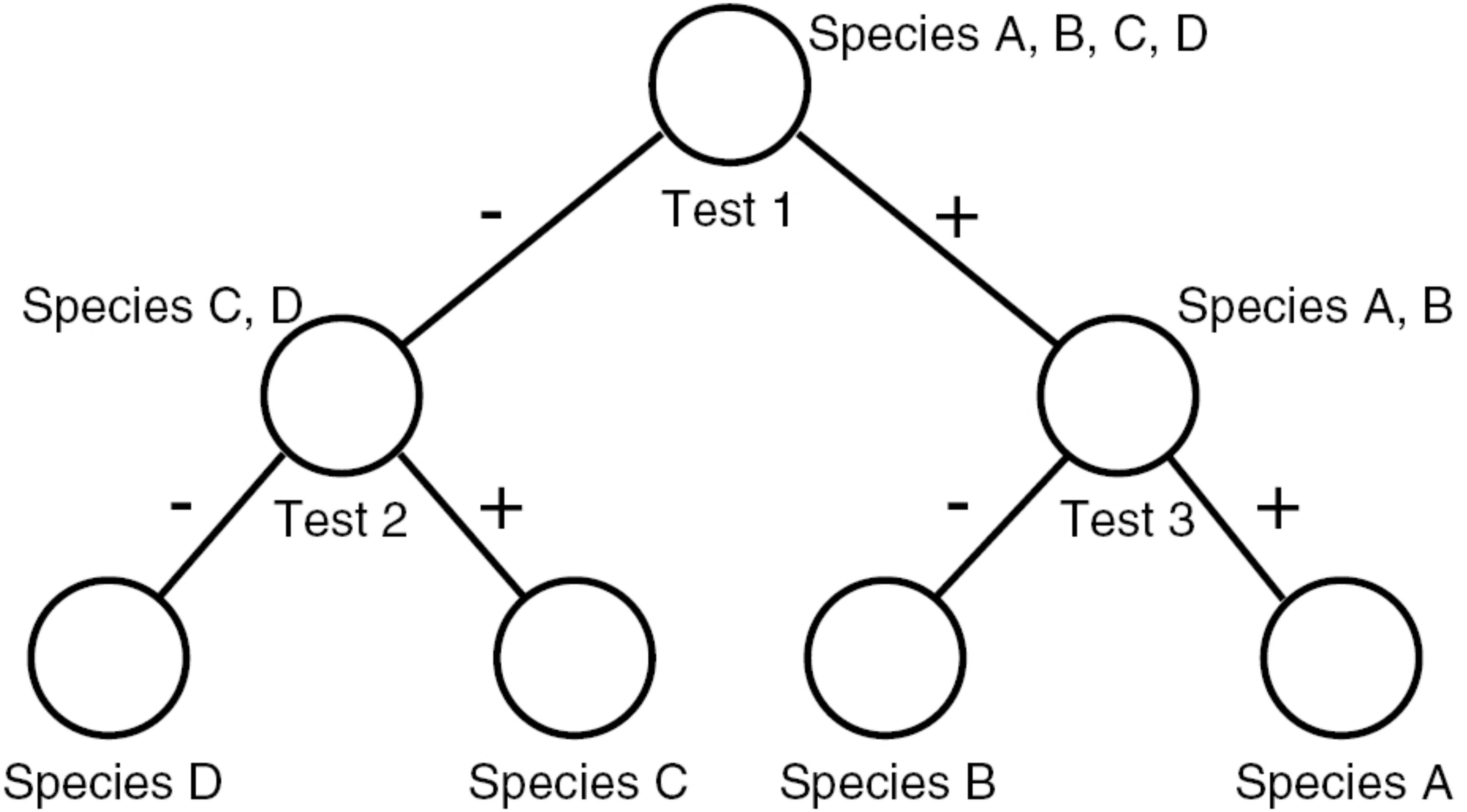
$$log2N$$

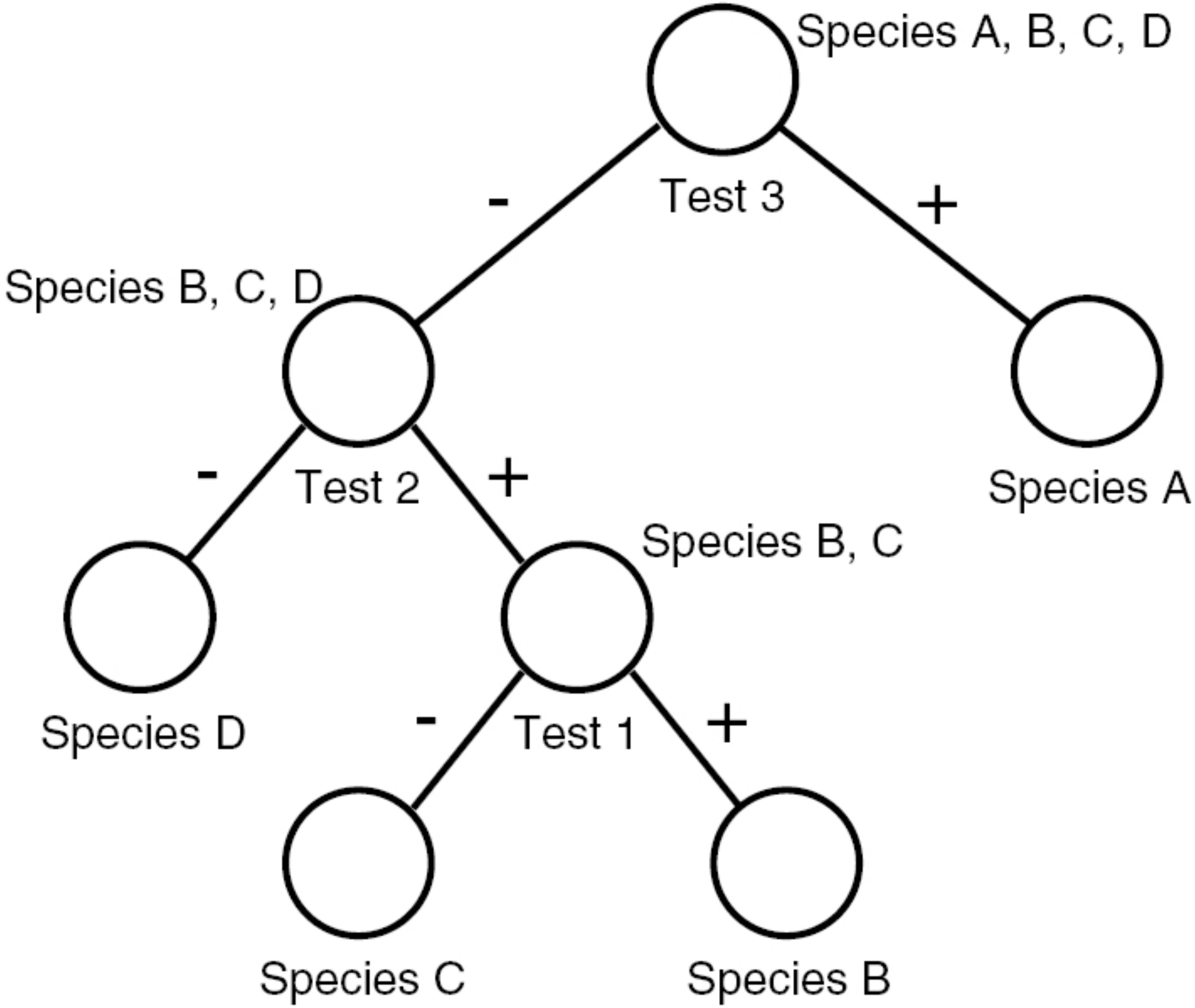
where N is the total number of organisms in the key.

The “worst” characters, on the contrary, are those separating only one organism. By using the “worst” 2-state characters only, a key will result

$$(N–1)(N+2)/(2N)$$

steps long.


Species A, B, C, D
Test 1
-
+
Species C, D
Species A, B
Test 2
-
+
Test 3
-
+
Species D
Species C
Species B
Species A


Species A, B, C, D
Test 3
-
+
Species B, C, D
Species A
Test 2
-
+
Species B, C
Species D
Test 1
-
+
Species C
Species B

An estimate of the ratio

$$\frac{\text{cost}}{\text{efficency}}$$

is often used to define the concept of “best character” in the field of industrial identification, as an example of yeasts.

The concept of “cost” can be well defined as far as laboratory tests are concerned, but it could be more difficult in other contexts where an identification is performed.

The concept of “best” character in building a “successful” identification key depends on what the term “successful” means

A “successful” key is the one with the shortest branches, or the most easy to use by target users?

There are not “best” character(s), or “best” sequences, but there can exist “better” characters, and sequences, than others in different contexts.

The context can be defined by a combination of several parameters, as:

- target users
- group of organisms
- equipment
- season

A taxonomists only, which knows a group of organisms, can choose the most effective sequence of characters to build a successful key.

Other approaches

Identification

DNA barcode

Require **limited human intervention** in the process

Can work also on a **single fragment** of a specimen

Require a relevant **skill level**

Not **portable** in the field, at the moment

The identification process leads to **an array of names (0-n)**

Identification

DNA barcoding employs sequence diversity in short, standardized gene regions to aid species identification and discovery in large assemblages of life. Ideally, one gene sequence would be used for all taxonomic groups.

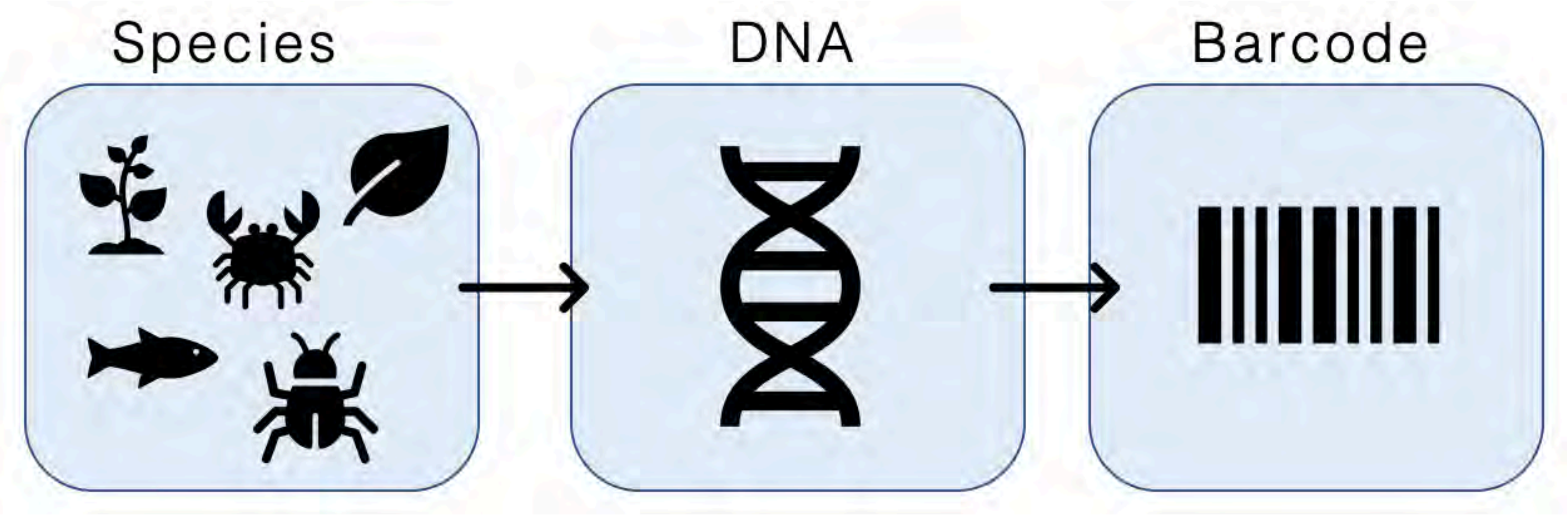
However, no suitable gene region has been found yet, so different barcodes are used for different groups of organisms.

For **animals**, the most widely used barcode is the mitochondrial cytochrome C oxidase I (COI) locus, while other mitochondrial genes are used as well. They are preferred over nuclear genes since they lack introns, are haploid, and have limited ricombination. Moreover, there are up to several thousand mitochondira in each cell.

In **plants** mitochondrial genes have a too low mutation rate. Thus, candidate genes have been found in the chloroplast genome, above all the maturase K (matK), often in association with other genes such as the ribosomial internal transcriber spacer (ITS), the ribulose-1,5-biphosphate carboxilase gene (rbcl), etc.

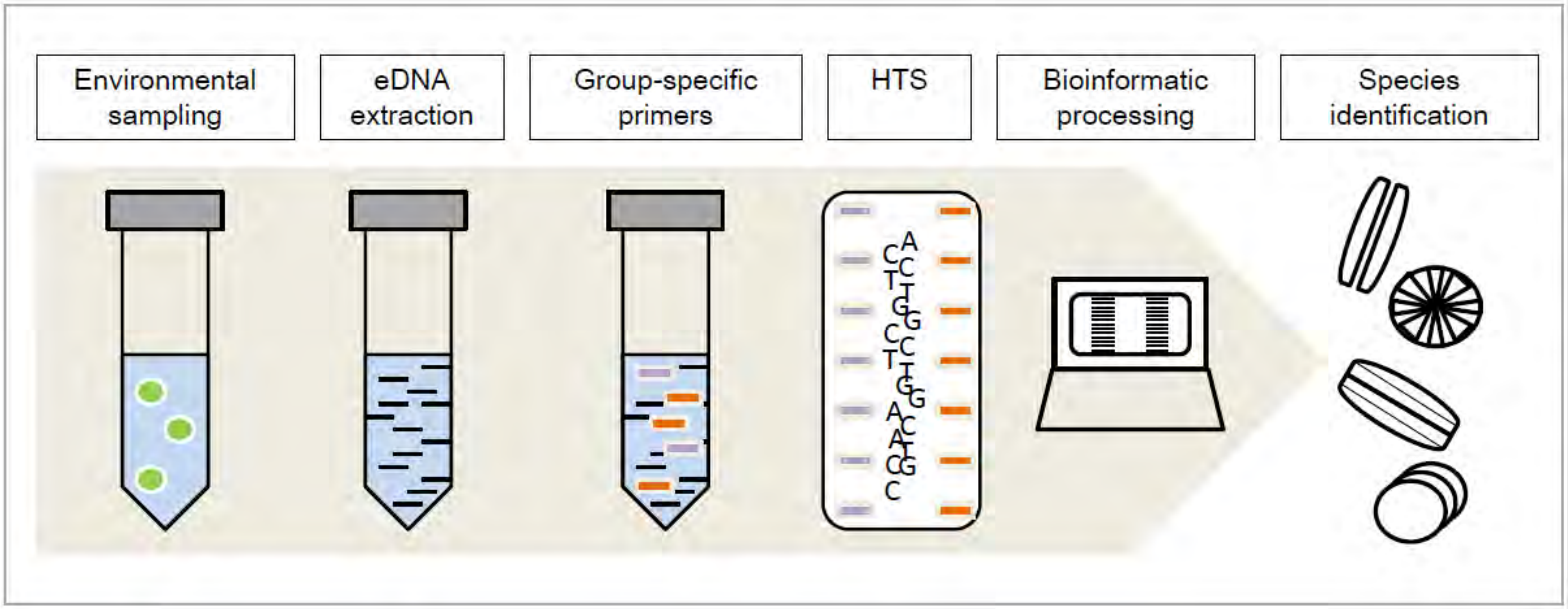
In **fungi** things are trickier. More than one locus has to be used at the same time. Normally, ITS rDNA, togfether with the large subunit of nuclear ribosomal RNA (LSU) are used.

Identification



Identification

However, the process can be used to face identification of organisms in environmental DNA samples.



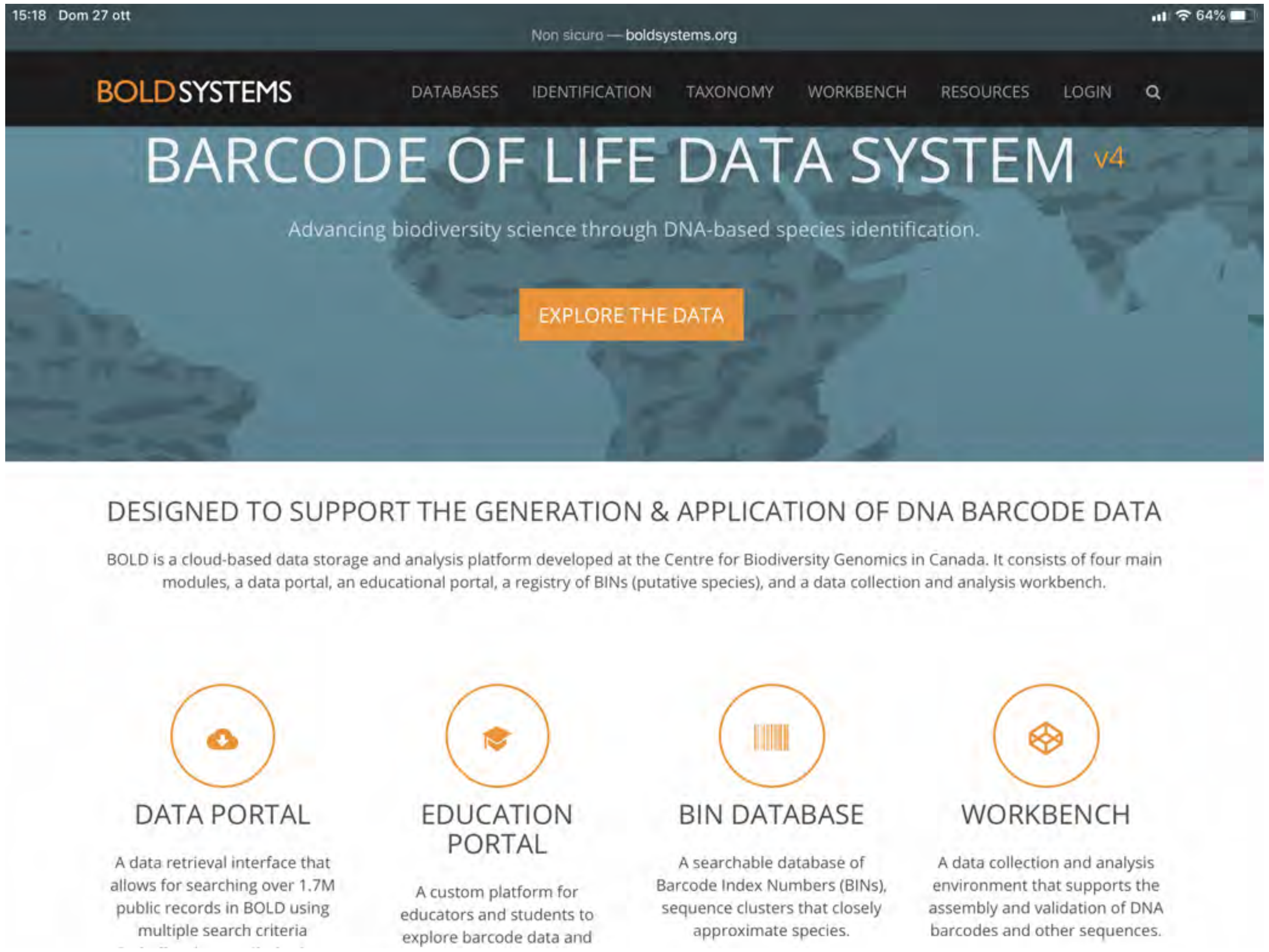
Identification

The Barcode of Life Data System (bold) — www.boldsystem.org — provides an integrated bioinformatics platform supporting all phases of the analytical pathway from specimen collection to tightly validated barcode library.

It is a repository for the specimen and sequence records that form the basic data unit of all barcode studies.

It is a workbench that aids the management, quality assurance and analysis of barcode data.

It provides also a vehicle for collaboration among diverse research centres.



BARCODE OF LIFE DATA SYSTEM v4

Advancing biodiversity science through DNA-based species identification.

DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODE DATA

BOLD is a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINs (putative species), and a data collection and analysis workbench.

DATA PORTAL

A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria

EDUCATION PORTAL

A custom platform for educators and students to explore barcode data and

BIN DATABASE

A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

WORKBENCH

A data collection and analysis environment that supports the assembly and validation of DNA barcodes and other sequences.

Identification

Automated image recognition

Require **limited human intervention** in the process

Do not require any **skill level**

Normally need a **“good” specimen** to perform at their best

Available on a **digital** support

Are **portable** in the field

Normally used **off-line**

The identification process leads
to **an array of names (0-n)**

Identification

Relevant characters for automated identification

Leaf shape A multitude of methods for its description can be found in literature. Also, most traditional taxonomic keys involve leaf shape for discrimination. While traditional identification categorizes leaf shape into classes (e.g., ovate, oblique, oblanceolate), computerized shape descriptors either analyze the contour or the whole region of a leaf. Initially, basic geometric descriptors, such as aspect ratio, rectangularity, circularity, and eccentricity, were used to describe a shape. Later, more sophisticated descriptions, such as center contour distance, Fourier descriptors, and invariant moments, were intensively studied.

The **vein structure**. Venation extraction is not trivial, mainly due to a possible low contrast between the venation and the rest of the leaf blade structure. Some authors have simplified the task by using special equipment and treatments that render images with more clearly identified veins. However, this defeats the goal of having users get an automated identification for specimens that they have photographed with ordinary digital cameras.

Identification

Relevant characters for automated identification

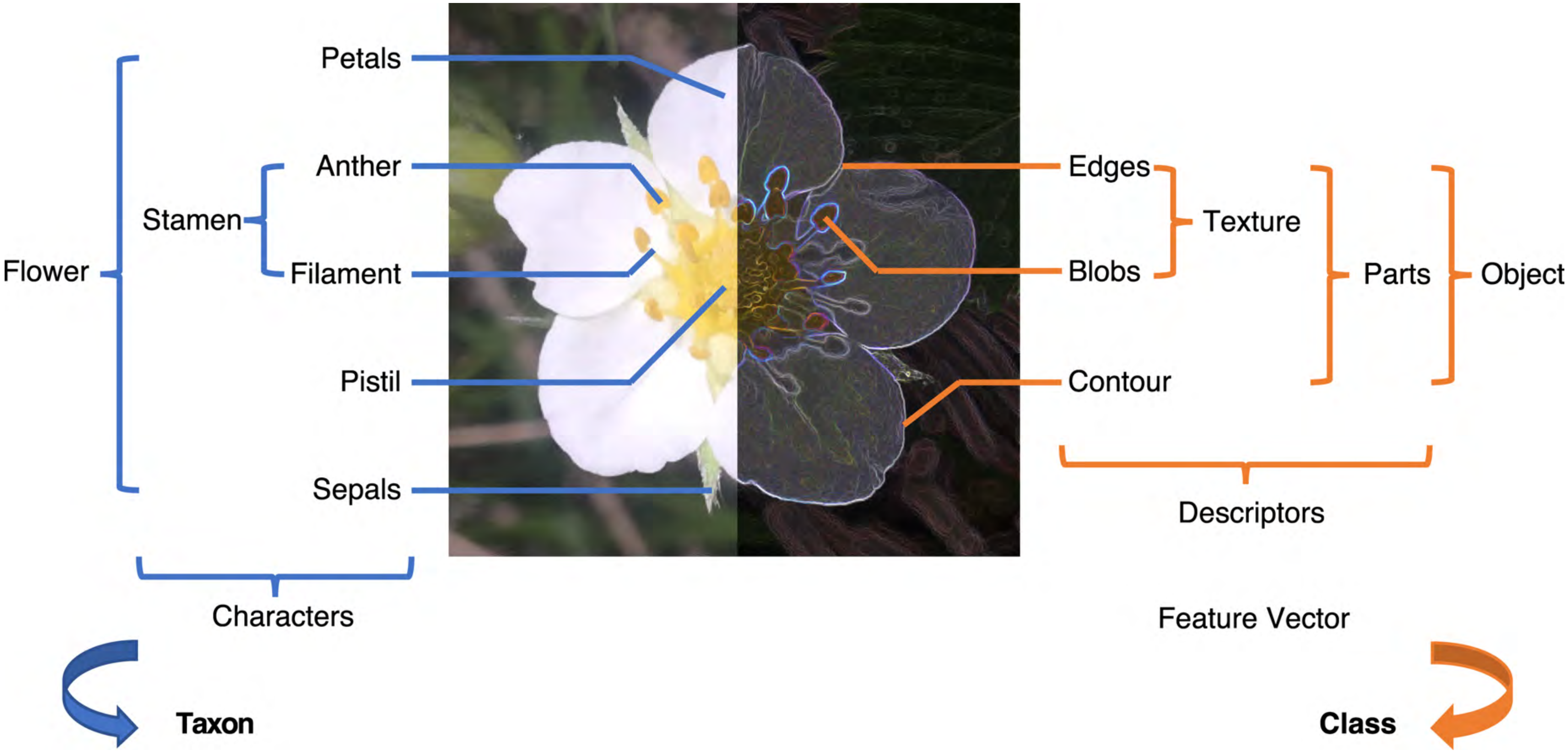
Leaf color is considered a less discriminative character than shape and texture.

Flower shape has hardly been considered so far. Interestingly, flower shape is an important characteristic in the traditional identification process. However, previous attempts for describing flower shape in a computable form did not find it to be very discriminative. A major reason is the complex 3D structure of flowers, and their changes in time.

Flower color is a more discriminative character. For automated identification, color has been mostly described by color moments and color histograms. Due to the low dimensionality and the low computational complexity of these descriptors, they are also suitable for real-time applications. However, solely analyzing color characters, without, e.g., considering flower shape, cannot classify flowers effectively.

Various previous studies showed that no single character may be sufficient to separate all desired taxa, making character selection and description a challenging problem.

How do we see a flower: botanist vs. computer



Flower

Stamen

Petals

Anther

Filament

Pistil

Sepals

Characters

Taxon

Edges

Blobs

Contour

Texture

Parts

Object

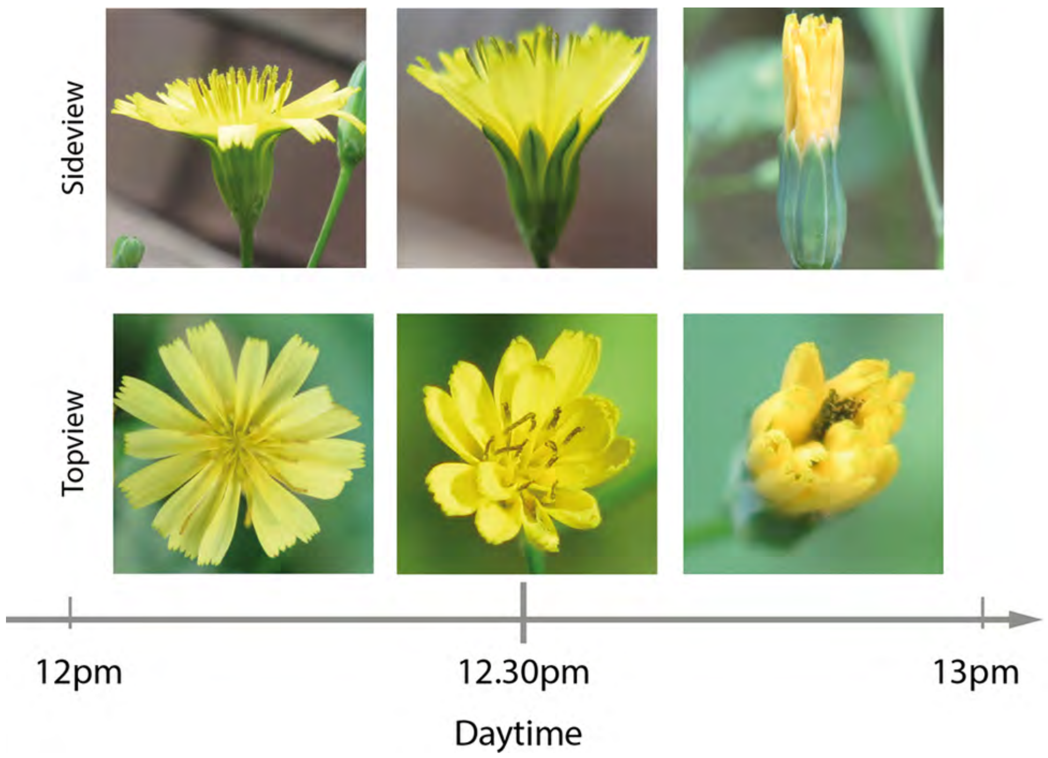
Descriptors

Feature Vector

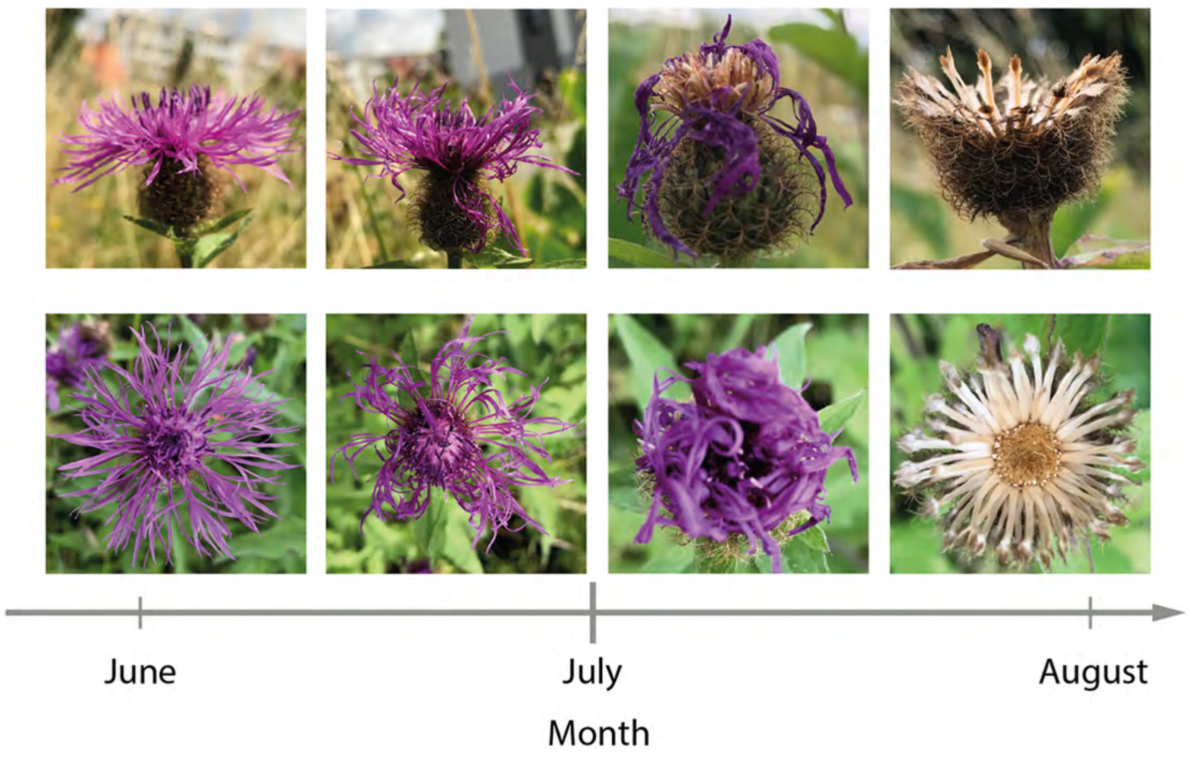
Class

Identification

Large intra-specific visual variation



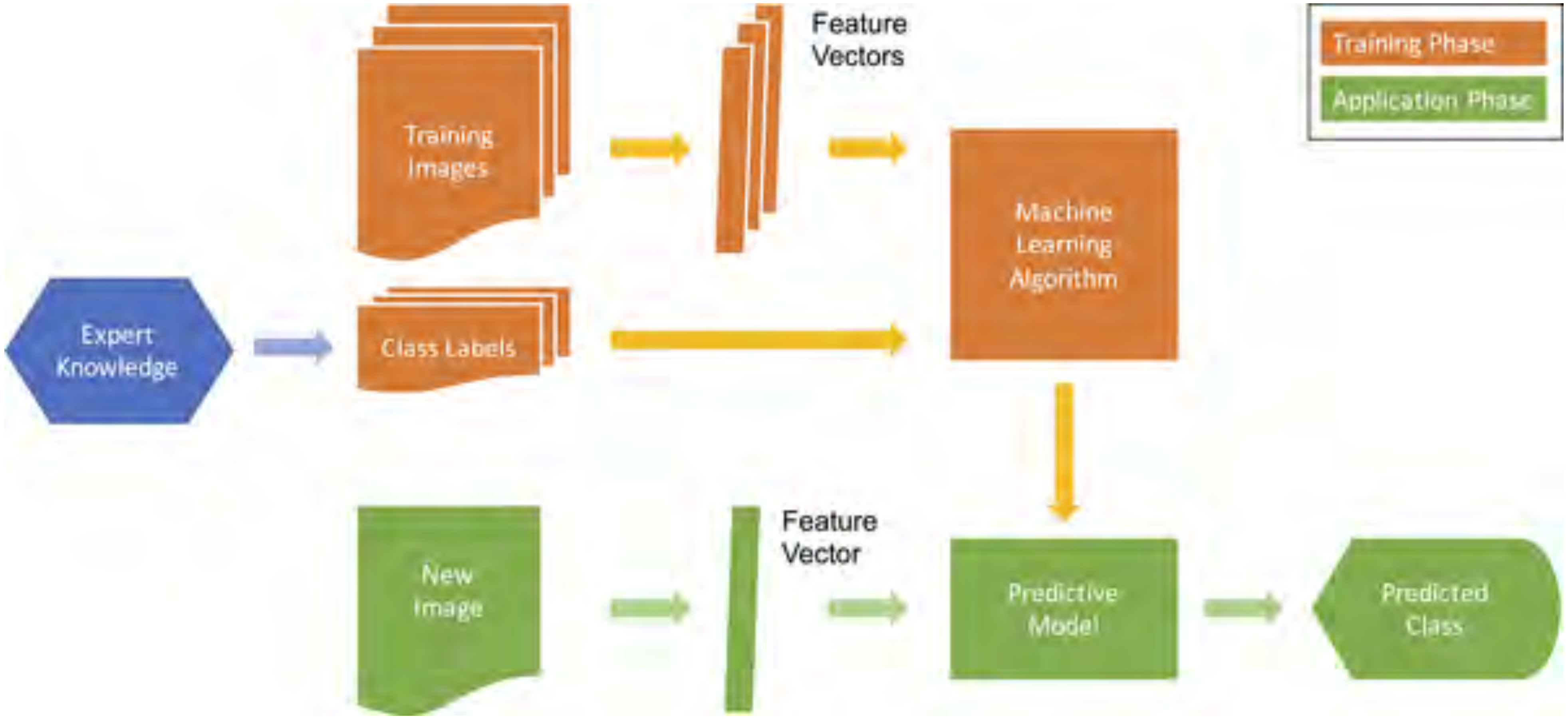
Lapsana communis



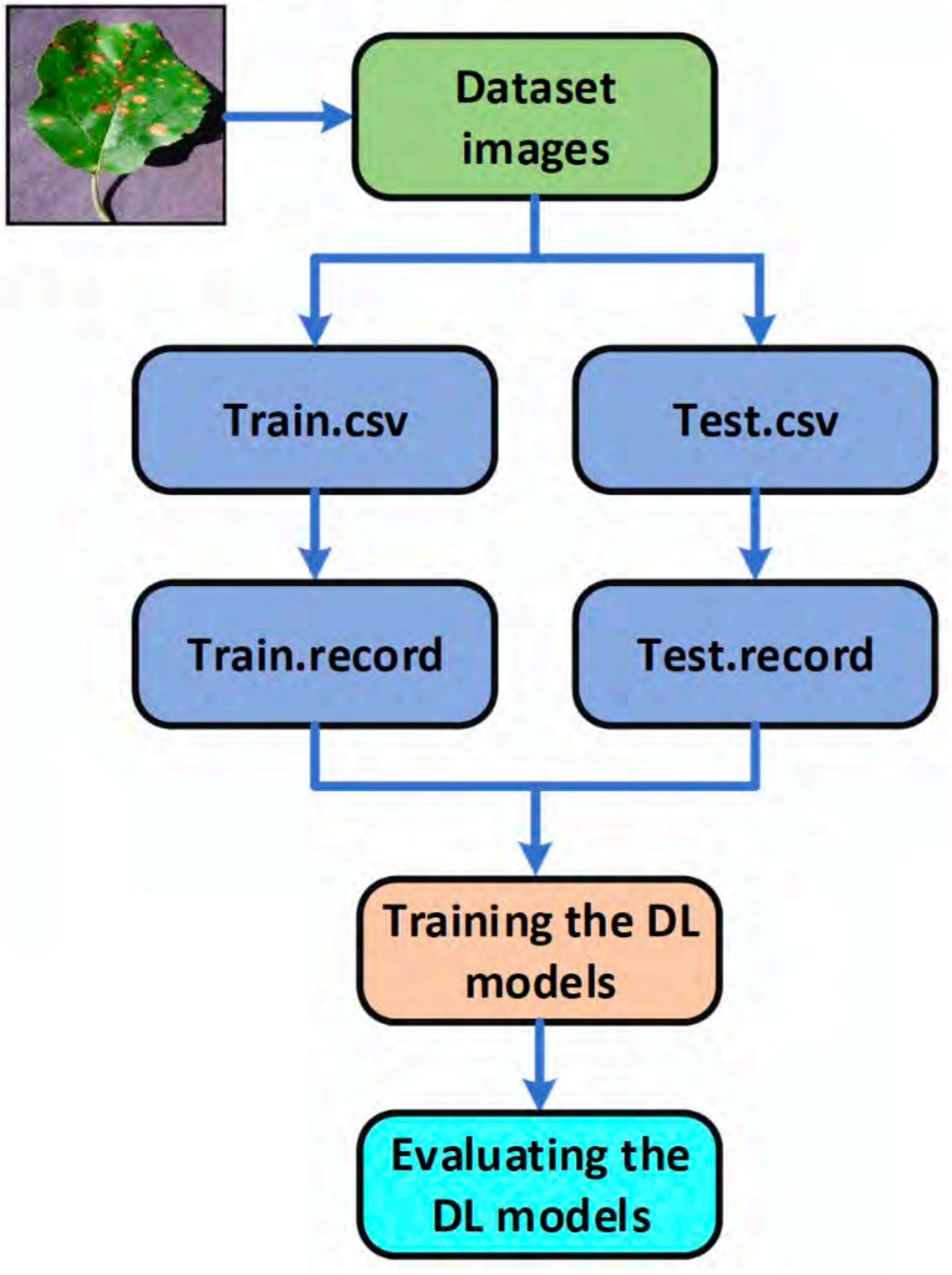
Centaurea pseudophrygia

Identification

Fundamental steps of supervised machine learning for image-based species identification.



Identification

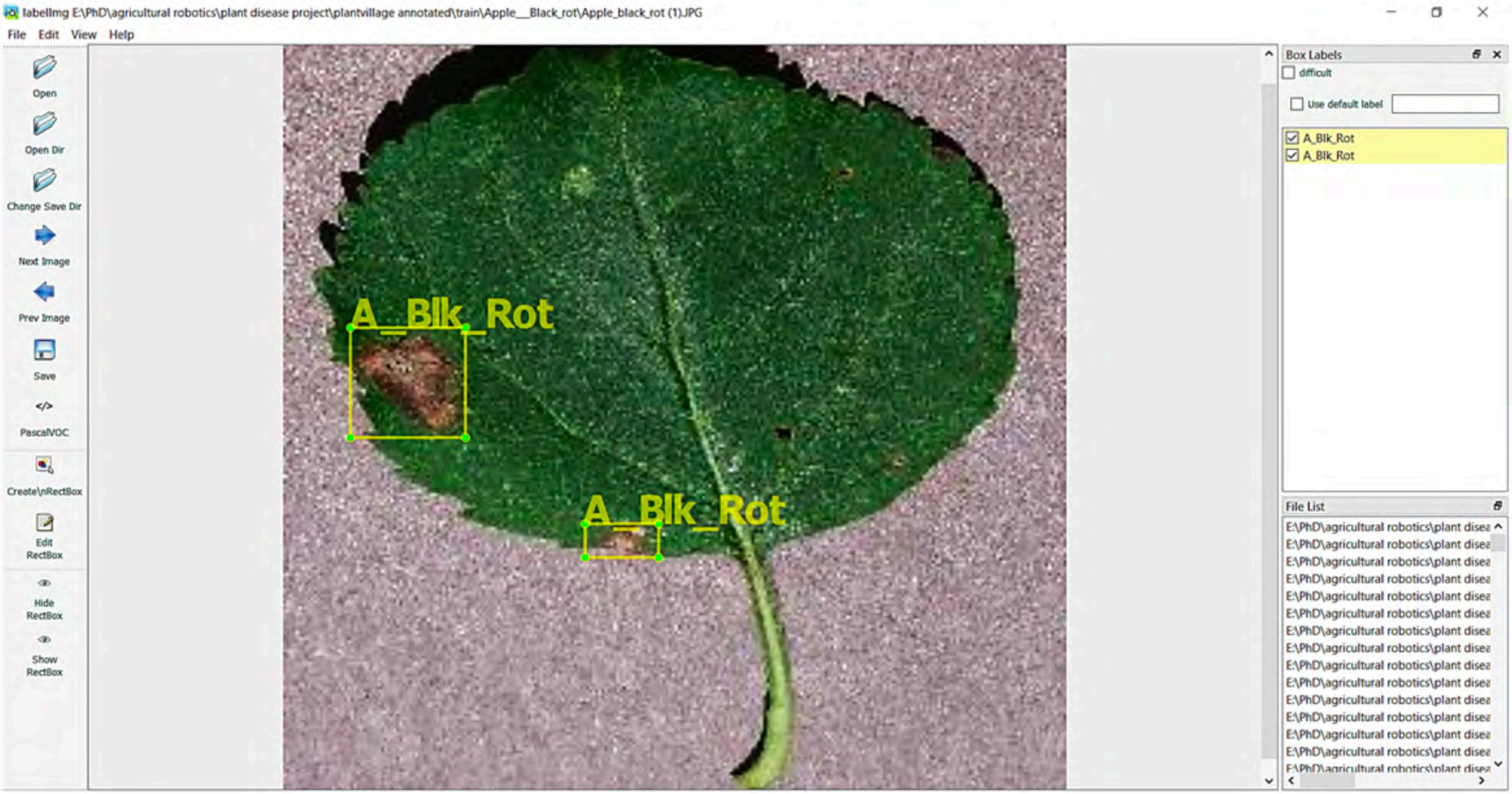


Plant diseases

Plant diseases are a major economic issue, especially in agricultural ecosystems. They need to be addresser properly, in order to avoid loss of money and time.
Thus, a correct identification of the pathogen is mandatory for control and mitigation.
Image recognition tools can support the identification of plants diseases on the basis of the morphology of the infected portions of the plants.

Certified images are used as a test set for training identification algorithms, which are then tested on other sets of images (test sets).

Identification



labelImg E:\PhD\agricultural robotics\plant disease project\plantvillage annotated\train\Apple___Black_rot\Apple_black_rot (1).JPG

☐ difficult

☐ Use default label

☑ A_Blk_Rot
☑ A_Blk_Rot

E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea
E:\PhD\agricultural robotics\plant disea

Identification

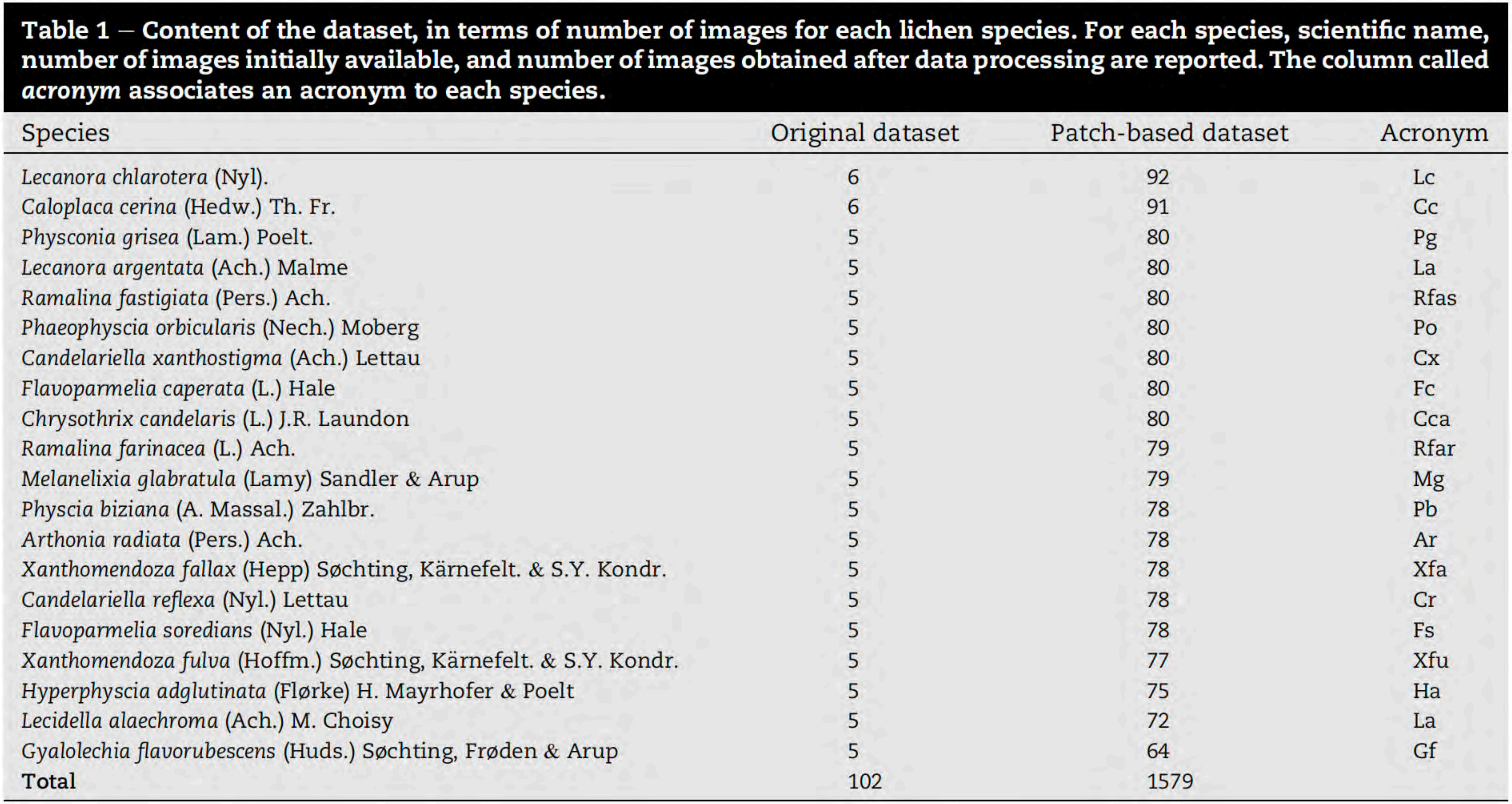
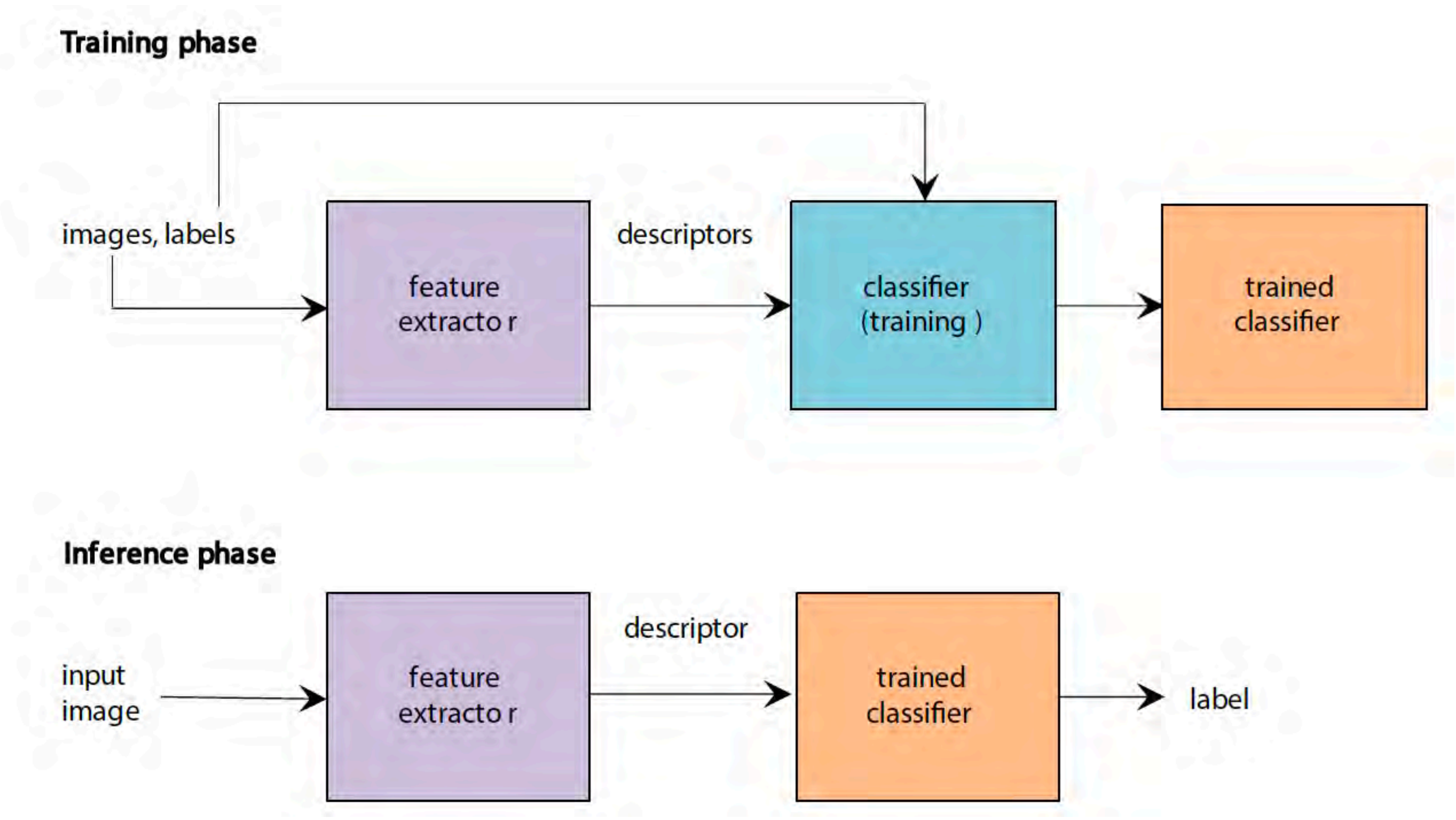


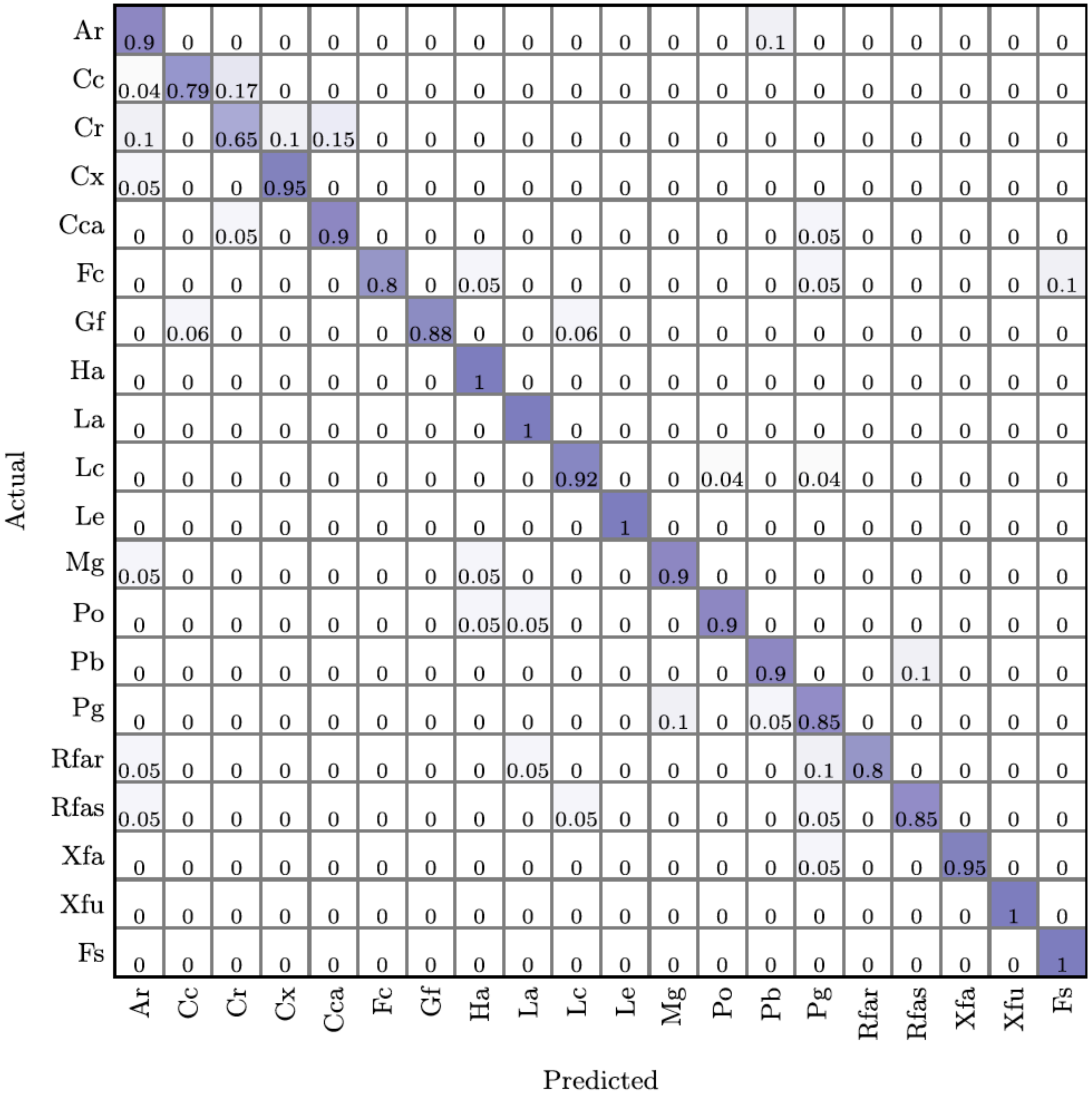
Table 1 – Content of the dataset, in terms of number of images for each lichen species. For each species, scientific name, number of images initially available, and number of images obtained after data processing are reported. The column called *acronym* associates an acronym to each species.

Species	Original dataset	Patch-based dataset	Acronym
Lecanora chlarotera (Nyl).	6	92	Lc
Caloplaca cerina (Hedw.) Th. Fr.	6	91	Cc
Physconia grisea (Lam.) Poelt.	5	80	Pg
Lecanora argentata (Ach.) Malme	5	80	La
Ramalina fastigiata (Pers.) Ach.	5	80	Rfas
Phaeophyscia orbicularis (Nech.) Moberg	5	80	Po
Candelariella xanthostigma (Ach.) Lettau	5	80	Cx
Flavoparmelia caperata (L.) Hale	5	80	Fc
Chrysothrix candelaris (L.) J.R. Laundon	5	80	Cca
Ramalina farinacea (L.) Ach.	5	79	Rfar
Melanelixia glabratula (Lamy) Sandler & Arup	5	79	Mg
Physcia biziana (A. Massal.) Zahlbr.	5	78	Pb
Arthonia radiata (Pers.) Ach.	5	78	Ar
Xanthomendoza fallax (Hepp) Søchting, Kärnefelt. & S.Y. Kondr.	5	78	Xfa
Candelariella reflexa (Nyl.) Lettau	5	78	Cr
Flavoparmelia soredians (Nyl.) Hale	5	78	Fs
Xanthomendoza fulva (Hoffm.) Søchting, Kärnefelt. & S.Y. Kondr.	5	77	Xfu
Hyperphyscia adglutinata (Flørke) H. Mayrhofer & Poelt	5	75	Ha
Lecidella alaechroma (Ach.) M. Choisy	5	72	La
Gyalolechia flavorubescens (Huds.) Søchting, Frøden & Arup	5	64	Gf
Total	102	1579	

Identification



Identification



	Ar	Cc	Cr	Cx	Cca	Fc	Gf	Ha	La	Lc	Le	Mg	Po	Pb	Pg	Rfar	Rfas	Xfa	Xfu	Fs
Ar	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0
Cc	0.04	0.79	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cr	0.1	0	0.65	0.1	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cx	0.05	0	0	0.95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cca	0	0	0.05	0	0.9	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0
Fc	0	0	0	0	0	0.8	0	0.05	0	0	0	0	0	0	0.05	0	0	0	0	0.1
Gf	0	0.06	0	0	0	0	0.88	0	0	0.06	0	0	0	0	0	0	0	0	0	0
Ha	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
La	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Lc	0	0	0	0	0	0	0	0	0	0.92	0	0	0.04	0	0.04	0	0	0	0	0
Le	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Mg	0.05	0	0	0	0	0	0	0.05	0	0	0	0.9	0	0	0	0	0	0	0	0
Po	0	0	0	0	0	0	0	0.05	0.05	0	0	0	0.9	0	0	0	0	0	0	0
Pb	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0.1	0	0	0
Pg	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.05	0.85	0	0	0	0	0
Rfar	0.05	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0.1	0.8	0	0	0	0
Rfas	0.05	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0.05	0	0.85	0	0	0
Xfa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0.95	0	0
Xfu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Fs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1


"That's all Folks!"