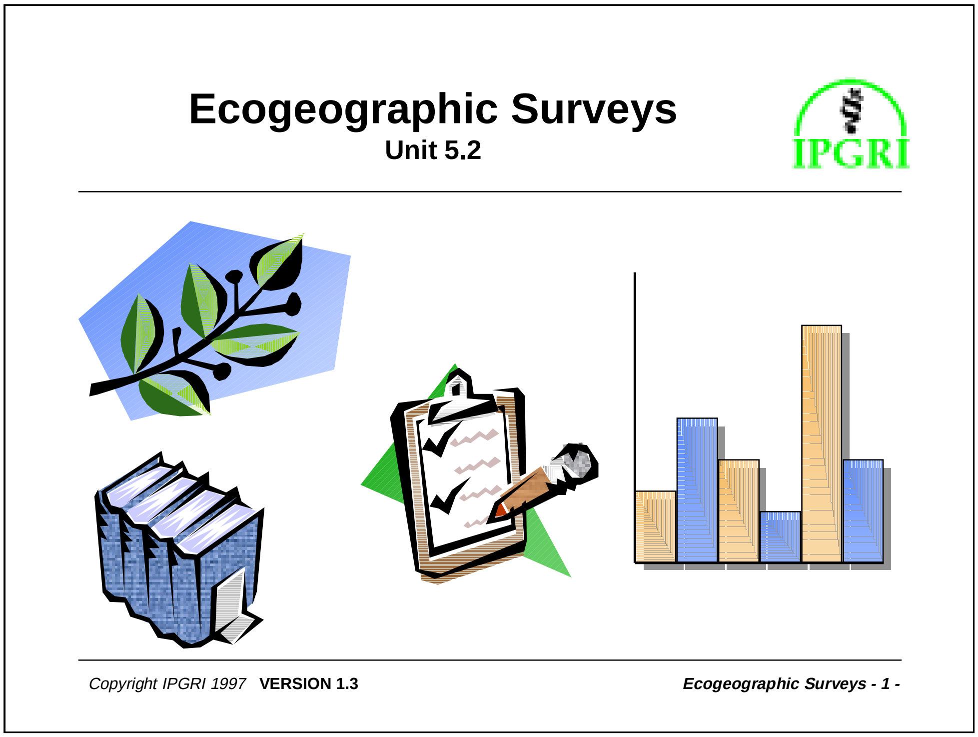


Unit 5.2

Ecogeographic Surveys

Version 1.3

Nigel Maxted, University of Birmingham

Kevin Painting, IPGRI, Rome

Luigi Guarino, IPGRI, Nairobi

Objectives

 VERSION 1.3

Ecogeographic Surveys - 2 -

Objectives

Upon completion of the unit the trainee will be able to:

- define ecogeographic surveys and give examples of their application in plant genetic resources work
- summarise the three different phases of an ecogeographic survey, listing the objectives, activities and outputs of each phase
- list the different types of geographical, ecological and taxonomic data which are collected in an ecogeographic survey
- identify problems commonly encountered when recording and analysing ecogeographic data and suggest ways of reducing them
- review the different statistical and graphical methods available for analysing and displaying ecogeographic data
- list the contents of an ecogeographic conspectus and report and summarise the criteria used in setting conservation priorities

Outline

IPGRI

- Introduction
- Phase 1 - Project Design
- Phase 2 - Data Collection and Analysis
- Phase 3 - Product Preparation
- Conclusion

 Ecogeographic Surveys - 3 -

Introduction

- **The Threat of Genetic Erosion**

Ecogeographic studies are introduced against the background of global genetic erosion.

- **A Definition**

Ecogeographic data are described. A distinction is made between ecogeographic studies and ecogeographic surveys.

- **Ecogeographic Surveys and Conservation**

The use of ecogeographic data in conservation activities is emphasised with an example given of the planning of collecting missions.

- **Ecogeographic Surveys: a Methodology**

The three major phases of an ecogeographic survey are outlined along with the three major outputs.

The Threat of Genetic Erosion

- Genetic erosion - a global issue
- The use of ecogeographic data
  - in identifying and assessing the threat of genetic erosion
  - in developing conservation strategies
  - in planning collecting missions

 Ecogeographic Surveys - 4 -

The current threat to plant genetic resources from global genetic erosion has led to an increased need for their conservation through *in situ* and *ex situ* conservation strategies. As discussed in Unit 8.1.1, genetic erosion can be caused by a number of factors including:

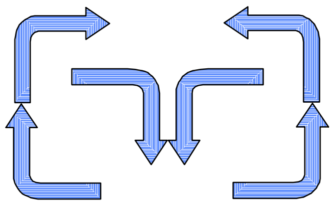
- agricultural change
- socioeconomic change
- overexploitation
- habitat loss
- competitors, predators and pests
- natural disasters/pollution.

To recognize and react to any threat of genetic erosion of plant species or populations, accurate, timely and relevant information is needed. Ecogeographic studies can provide invaluable information in assessing the threat of genetic erosion. Ecogeographic information can be used for the identification of conservation priorities and the development of conservation strategies. Importantly, it is essential for planning collecting missions.

A Definition

"An ecogeographic study is a process of gathering and synthesizing ecological, geographical and taxonomic information. The results are predictive and can be used to help formulate conservation strategies and collecting priorities."

Maxted et al. *(1995*)



 Ecogeographic Surveys - 5 -

Ecogeographic data comprise ecological, geographic and taxonomic data and the data sets are often large and complex. Data can be obtained from the literature and/or from the compilation of passport data from herbarium specimens and germplasm accessions.

An **ecogeographic study** uses considerable resources to carry out and may take several years to complete. In ecogeographic studies, it is common to gather data by sampling directly the target species or population. The analysis of the ecogeographic data collected is usually very detailed.

Example: Ecogeographic study by Ehrman & Cocks (1990) for the annual legumes of Syria.

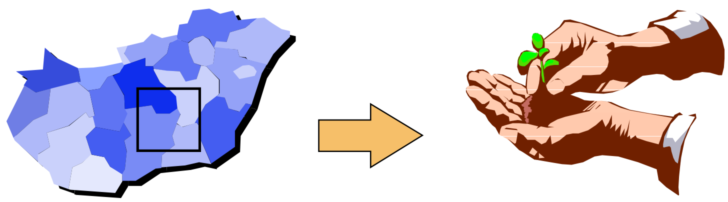
From the very detailed ecogeographic data gathered over several years by the authors and by extensive analyses of climatic and soil characteristics that influenced the distribution of the annual legume species, they found that species diversity and seed production were related to annual rainfall and that populations in the drier areas faced a greater threat of genetic erosion. Following the analysis, they proposed a detailed list of conservation priorities.

An **ecogeographic survey** tends to rely on data recorded by other plant collectors rather than obtaining new data. It may, for instance, be limited to collating data from herbarium specimens and genebank accessions and performing a literature search.

Ecogeographic Surveys and Conservation

Example - Planning Collecting Missions

- To identify target species, collecting areas and habitats
- To predict where a species may be located



All conservation activities involve the collection and analysis of ecogeographic data.

Example: Planning Collecting Missions

Ecogeographic surveys or studies are particularly useful when planning collecting missions:

- **To identify target species/populations, collecting areas and habitats**

Resources available for germplasm collecting are always limited and as a result the most efficient use must be made of them. Therefore, each collecting mission should have a clearly defined set of target taxa, target areas and target habitats. One way to define these targets is to undertake an ecogeographic study or survey prior to the mission. This is particularly true for wild species as much time can be wasted if the habitat preferences or geographical distribution of the target species are not known prior to the start of the collecting mission.

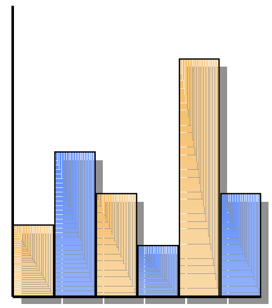
- **To predict where a species may be located**

The locations inhabited by each plant species or population are defined by particular sets of environmental and geographical conditions. By knowing what these conditions are, one can predict where a species may be located. The ecogeographic passport data associated with herbarium specimens and germplasm accessions can be used to decide what the set of environmental conditions are likely to be. Moreover, a combination of ecological and geographical passport data from historical collections provides evidence that can be used to predict where species may be currently located.

Ecogeographic Surveys: a Methodology

IPGRI

- Phases
  - Project Design
  - Data Collection and Analysis
  - Product Preparation
- Outputs
  - Database
  - Conspectus
  - Report



 Ecogeographic Surveys - 7 -

There are three major phases involved in an ecogeographic survey or study which can be summarised as follows:

Phase 1: Project Design

- project commissioning, selecting and delimiting the target taxonomy and area, identifying suitable taxonomists and taxonomic collections

Phase 2: Data Collection and Analysis

- survey of geographical, ecological and taxonomic data, collection of specimen specific data, data verification and analysis

Phase 3: Product Preparation

- production of the ecogeographic database, conspectus and report.

The information gathered and generated as a result of the study are presented in three major outputs, namely:

- **The ecogeographic database** - containing the raw data for each taxon
- **The ecogeographic conspectus** - summarising the data for each taxon
- **The ecogeographic report** - discussing the contents of the database and proposing future conservation strategies and collecting missions as appropriate.

Outline

- Introduction
- Phase 1 - Project Design
- Phase 2 - Data Collection and Analysis
- Phase 3 - Product Preparation
- Conclusion

 Ecogeographic Surveys - 8 -

Phase 1 - Project Design

▪ **Project Commissioning**

The taxonomic and geographic scope of the project is defined and an appropriate specialist selected to do the work.

▪ **Identification of Suitable Taxonomists**

Taxonomic specialists are sought who can give advice on taxonomic matters in the survey.

▪ **Selection of the Target Taxonomy**

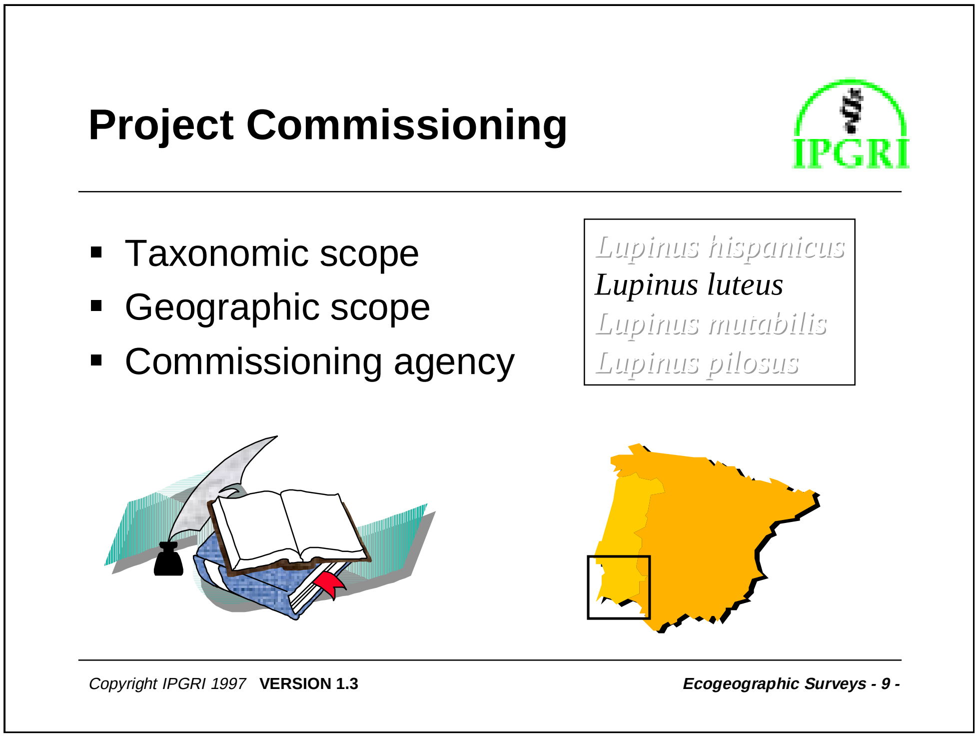
The most appropriate classification system for the target taxon is chosen and further sources of data on the taxon sought.

▪ **Delimitation of the Target Area**

Careful consideration is given to the region(s) to be included in the survey.

▪ **Identification of Taxonomic Collections**

The most appropriate herbaria and genebanks are identified for visiting during the survey.



In order to justify conducting an ecogeographic survey, the target species must be considered of sufficient interest or value for the survey and subsequent conservation and use. Accordingly, the scope of the survey is restricted taxonomically.

Unless the target species is being studied throughout its range (as in ecogeographic studies), the target area is also likely to be delimited by resource limitations and the time available.

The commissioning agency can range from international conservation agencies (such as the Botanic Gardens Conservations Secretariat, World Wildlife Fund, International Plant Genetic Resources Institute, etc.) to a research institute sponsoring a collecting mission.

Project Commissioning - an Example

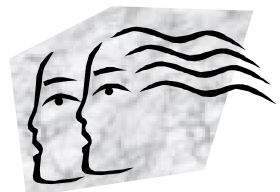
*"A general survey of **Corchorus** L. species was commissioned by the International Jute Organisation to provide the necessary background data on which future germplasm collecting expeditions could be based ... the survey was required to identify those wild species for potential use in the future genetic improvement of jute, in addition to identifying the countries and locations where collecting expeditions would be most profitable ..."* Edmonds (1990)

 Ecogeographic Surveys - 10 -

An example of a project commission is provided by Edmonds (1990) in her herbarium survey of African *Corchorus* species: "A general survey of *Corchorus* L. species was commissioned by the International Jute Organisation to provide the necessary background data on which future germplasm collecting expeditions could be based ... the survey was required to identify those wild species for potential use in the future genetic improvement of jute, in addition to identifying the countries and locations where collecting expeditions would be most profitable."

Project Commissioning - Selection of a Specialist

- Desirable qualities
  - familiarity with ecogeographic data
  - familiarity with the target group
  - familiarity with the geography of target area



 VERSION 1.3 ***Ecogeographic Surveys - 11 -***

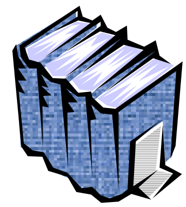
It is important that all data collected during an ecogeographic survey be relevant, accurate and reliable. Otherwise the subsequent predictions and conclusions drawn from the data could be misleading or wrong. Therefore, the survey should be undertaken or supervised by an appropriate specialist who is skilled in searching for, interpreting, recording and analysing data.

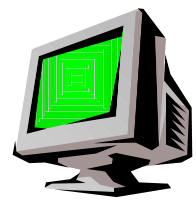
The specialist should have some background knowledge of the target species or group and be experienced in the use of identification aids since the taxonomy of wild species is sometimes problematic and the identification aids are sometimes of poor quality or even unavailable.

Familiarity with the geography of the target area is also desirable and can be useful when trying to interpret or decipher handwritten details, for instance when reading locality details on herbarium labels.

Identification of Suitable Taxonomists

- A taxonomist can assist with
  - locating relevant literature
  - identifying herbaria to visit
  - identifying other taxonomists
- Lists of taxonomists can be found
  - in *Index Herbariorum*
  - on the Internet





 VERSION 1.3 ***Ecogeographic Surveys - 12 -***

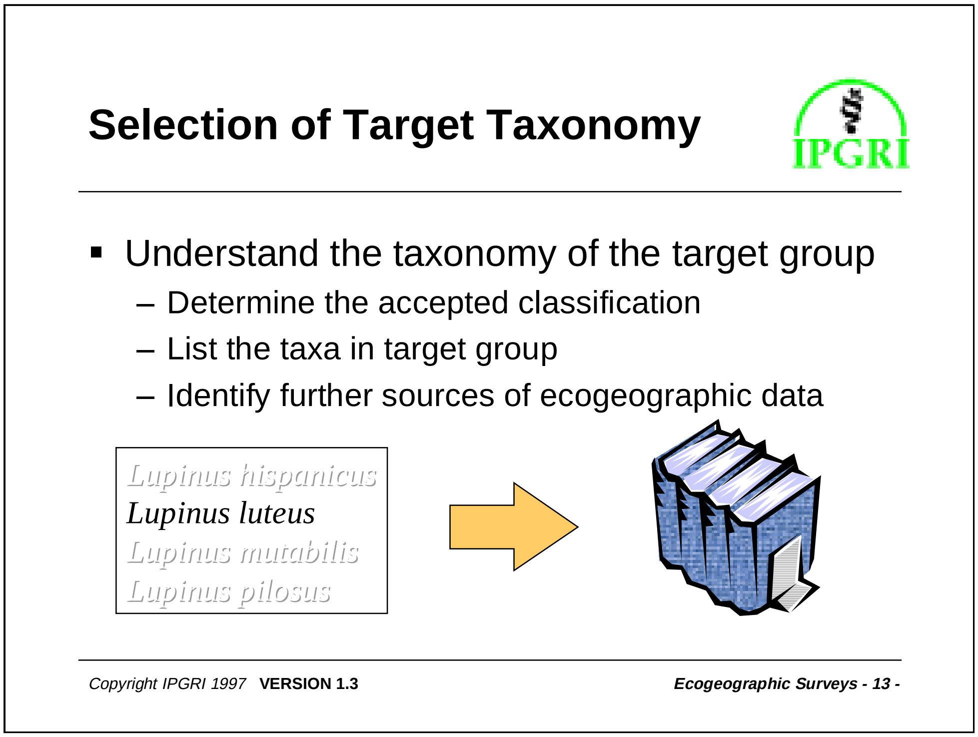
It is very useful to seek the advice of taxonomic specialists who can assist in the acquisition of ecogeographic data. Such specialists can give advice on:

- relevant literature, local floras and monographs
- herbaria and genebanks to visit
- other taxonomic specialists to consult.

Index Herbariorum (Holmgren *et al.* 1990) lists thousands of specialists worldwide, detailing their specialist taxonomic group and geographic region.

On the Internet, a World Wide Web site maintained by the University of Helsinki lists nearly 3,000 links to botanical resources available over the Internet. The subject category listing is updated weekly (Lampinen *et al.* 1996) and can be found at: http://www.helsinki.fi/kmus/botmenu.html

Contact name: Raino Lampinen, Email: Raino.Lampinen@helsinki.fi



It is crucial to have a good understanding of the taxonomy of the target group so that the ecogeographic data collected can be located and then meaningfully recorded, interpreted and analysed.

▪ **Determine the accepted classification of the target group**

A number of sources are used here: taxonomic specialists, Floras, monographs, recent revisions of the group, and various taxonomic studies which use both morphological and other types of data (e.g. molecular data). These will help determine the currently accepted classification of the group.

If the purpose of the survey is ultimately to conserve the maximum genetic variation in the target group, then a classification system based on the biological species concept may be more appropriate. This is because, traditionally, classification systems have concentrated on morphological characteristics which do not necessarily fully reflect the degree of genetic differentiation between the species. However, biologically based classification systems tend to be restricted to crop plants and their relatives where the make-up of the genepool is well understood and the genetic relationships among the taxa intensively studied.

▪ **List the target taxa and identify further sources of ecogeographic data**

The classification chosen will delimit the study by listing the taxa in the target group. It will also provide leads to other sources of data such as taxon descriptions, synonym lists, distribution maps, identification aids, ecological studies, bibliographies and critical taxonomic notes.

Delimitation of the Target Area

- Area should be chosen carefully
  - to maximise the predictive value of the survey
- Taxon should be studied
  - throughout its range ***or***
  - throughout a well-defined area, biogeographically or floristically

 VERSION 1.3 ***Ecogeographic Surveys - 14 -***

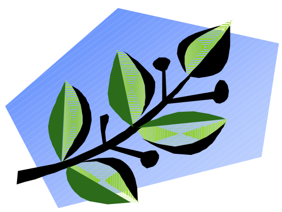
It is important that the target area is chosen carefully. This is to maximise the predictive value of the survey. Restricting the target area to save resources may lead, in the long term, to multiple studies of the same taxon being carried out, even by different researchers. This is likely to lead to a less coherent overall picture and will limit the predictive power of the studies.

Ideally, the specified taxon should be studied throughout its range but where this is not possible it should be studied throughout a well-defined area, biogeographically or floristically.

Once the target area is established, more information can be sought about the target taxon from local floras. These can give more detailed information on local geographic distribution or ecological preferences.

Identification of Taxonomic Collections

- Need to visit relevant herbaria and collections
  - – major sources of ecogeographic data
- Money and time restrictions
  - – choose most appropriate herbaria/collections



 VERSION 1.3 ***Ecogeographic Surveys - 15 -***

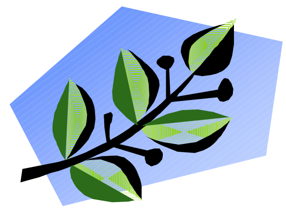
- **Choose herbaria and germplasm collections**

The researcher will need to visit the major herbaria and germplasm collections of the target taxon from the target area. This is because they are important sources of relevant ecogeographic data in the form of passport data from previous collecting missions.

Money and time restrictions usually limit the number of herbaria or collections that can be visited so it is important that the most appropriate ones be chosen. Taxonomists and area specialists can assist here in making the choice. *Index Herbariorum* also lists where the important herbarium collections are housed.

Herbaria: International or Regional?

- Aim to visit both types of herbaria
  - contain different specimens
  - availability of different expertise
  - broader sampling of data
- International herbaria
  - broad coverage
  - other resources (e.g. specialists, libraries)
  - predominance of older collections



It is advisable to visit **both** types of herbarium during the survey in view of the different specimens they contain and the availability of taxonomic or geographical specialists who can provide advice. Also, the broader the sampling of ecogeographic data associated with herbarium specimens and germplasm, the more likely the data will be predictive.

International Herbaria

Advantages

- Broad taxonomic coverage, possibly including material used in the production of revisions and monographs
- Broad international geographical coverage, possibly including material used in the production of local floras
- Skilled researchers available to provide general advice
- Appropriate taxonomic and geographical specialists
- Type material of target taxa
- Good botanical library.

Disadvantages

- Predominance of old collections, making extraction of passport data more difficult and reducing the likely predictive value of the data
- Geographical names associated with older collections sites may have changed more recently.

Herbaria: International or Regional?

- Regional herbaria
  - good local coverage
  - better documentation
  - regional specialists
  - fewer resources

 Ecogeographic Surveys - 17 -

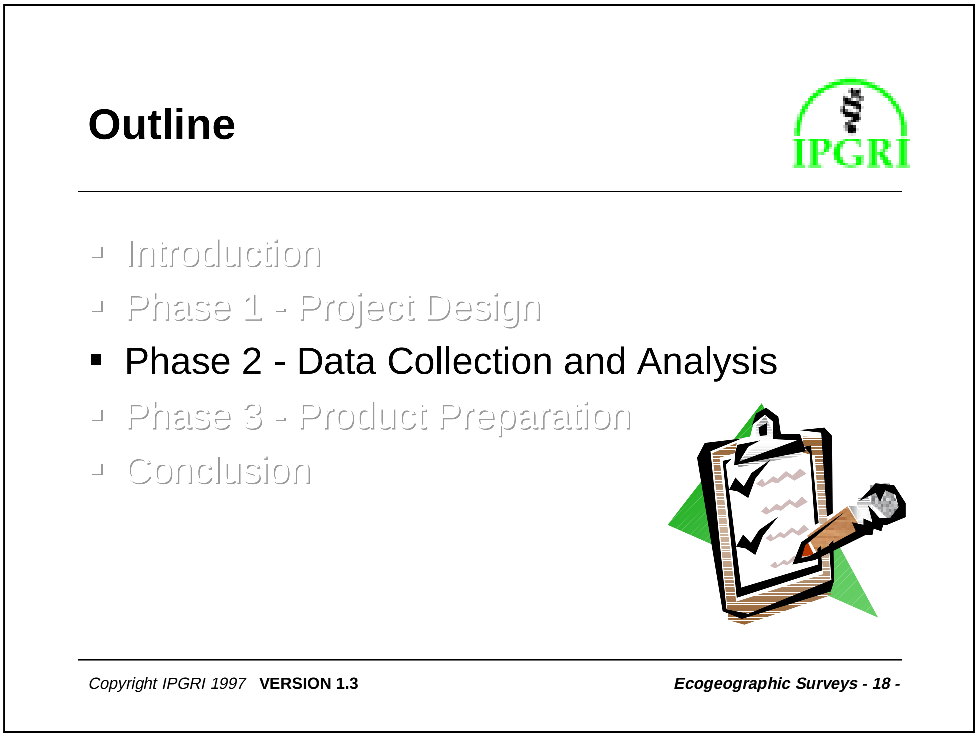
Regional Herbaria

Advantages

- Good local coverage of target region
- Better-documented material - herbarium is likely to have been more recently established
- Regional specialists present who can assist in deciphering local geographic names.

Disadvantages

- Limited resources for herbarium maintenance
- Lack of target taxon specialists
- Limited botanical library.



Phase 2 - Data Collection and Analysis

- **Assess current conservation activities**

The conservation status (*in situ* and *ex situ*) of the taxon for the target region is assessed.

- **Survey of geographical, ecological and taxonomic data**

Through a thorough survey of geographical, ecological and taxonomic data, the target area (geographic, climatic, etc.) is fully characterised along with the target species (distribution, ethnobotany, phenology, etc.).

- **Collection of specimen-specific data**

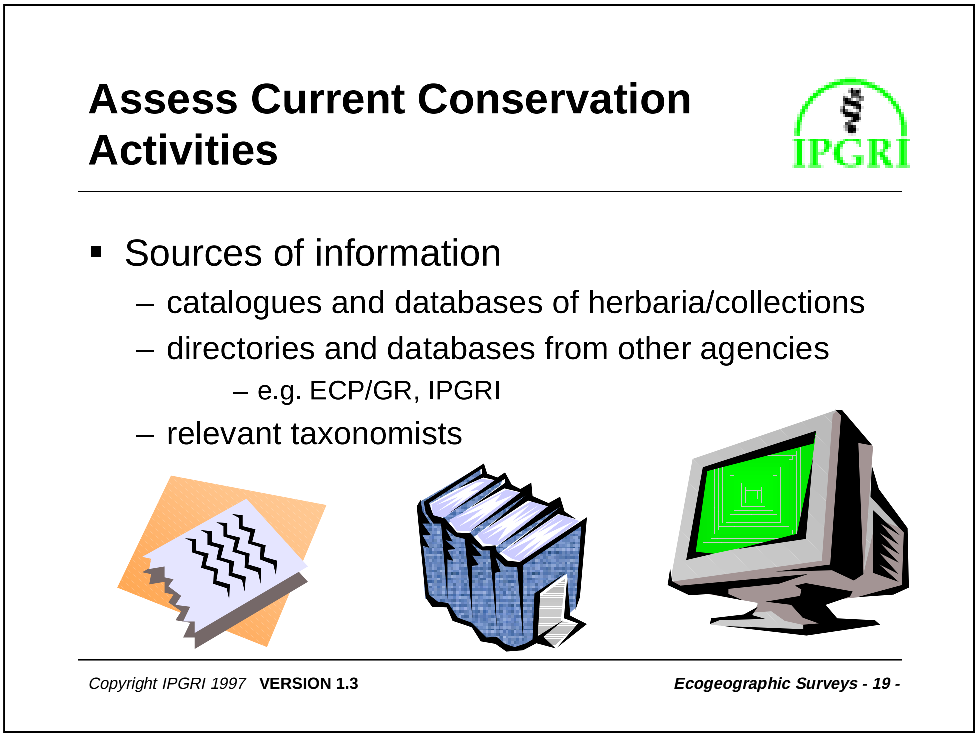
Specimen-specific data are collected in order to determine the distribution and ecological preferences of the target species. Considerations for data selecting and recording are given together with problems commonly encountered.

- **Data verification**

Data are verified for accuracy and integrity prior to analysis.

- **Data analysis**

Data analysis is performed to elucidate the distribution and ecological preferences of the target species. The application of maps, tables and bar charts, multivariate statistical analysis and Geographic Information Systems is discussed.



It is necessary to review the conservation status (*in situ* and *ex situ*) of the taxon from the target region, particularly if the purpose of the survey is to develop a conservation strategy. For instance, there may be little justification for collecting in the target region if the germplasm is already securely conserved in *ex situ* or *in situ* collections.

Details of what material is currently being conserved can be obtained from catalogues and databases from the respective botanical garden, germplasm collection and *in situ* conservation area. In addition, directories and databases of conserved material have been established by a variety of national and international agencies (e.g. European Cooperative Programme for Crop Genetic Resources Networks [ECP/GR], IPGRI). Taxon specialists can assist here in locating sources of information. Much of this information can be gathered in advance of any visits to herbaria and genebanks.

Problems in Interpreting Data (1)

- Conserved material is incorrectly identified
- Quantity of material conserved is misleading
- Availability of germplasm is restricted
- Status of material conserved

- *Conserved material is incorrectly identified*

This can only be verified by consulting voucher material or re-identifying living material, which can necessitate a visit to the collection. Clearly one can be confident of an indentification if it has been carried out by an expert for the taxonomic group.

- *Quantity of material conserved is misleading*

Genebanks and botanical gardens can duplicate their holdings in other collections around the world - this can give a misleading idea of the genetic diversity conserved.

- *Availability of germplasm is restricted*

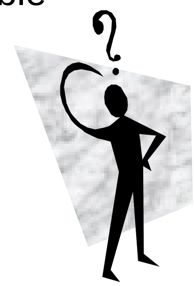
The germplasm conserved may be unavailable to the potential user because of restrictions on its supply or because it is conserved in only small quantities. This is not always clear from catalogues which could therefore give a false impression of the conservation status of the taxon.

- *The status of the material conserved*

The target population of a landrace or wild species will have more intrinsic variation than an advanced cultivar. It is therefore important to know the status of the sample conserved (e.g. breeder’s line, landrace, advanced cultivar etc.)

Problems in Interpreting Data (2)

- Genetic variation of original population is not represented in the samples
- Few or no passport data are available



 VERSION 1.3 ***Ecogeographic Surveys - 21 -***

- *Genetic variation of original population is not represented in the samples*
    - because the material was not sampled effectively
    - because of genetic drift during conservation
        - unsuitable regeneration conditions
        - poor storage resulting in differential erosion
        - human error e.g. mis-labelling of samples
- *Few passport data are available*

The value of the material for utilisation is limited where few or no passport data are available. This occurs more frequently with older germplasm collections.

Survey of Geographical, Ecological and Taxonomic Data

IPGRI

Taxon-specific data

- Accepted taxon name
- Locally used taxon name
- Where species grows
- Timing of local flowering/fruiting
- Habitat preference
- Topographic preference
- Soil preference
- Geological preference
- Climatic and microclimatic preference
- Breeding system
- Genotypic/phenotypic variation
- Biotic interactions
- Archaeological information
- Ethnobotanical information
- Conservation status

 Ecogeographic Surveys - 22 -

The collation of much of the geographical, ecological and taxonomic data can be carried out at the major herbaria as these usually have good botanical libraries. The data collected will be of two types, taxon specific and area specific.

Taxon-specific data

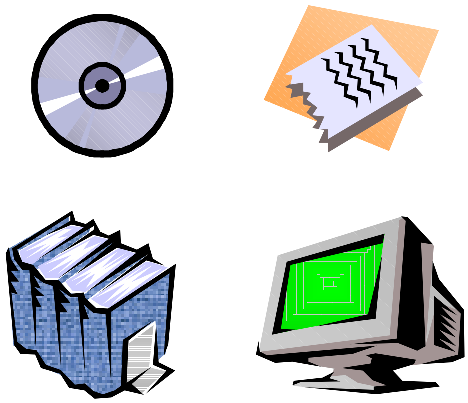
- Classification, distribution, reproduction, ethnobotany and phenology.

Together with specimen-specific data, these data can be used to ascertain the distribution and ecological preferences of the taxon.

- Accepted taxon name
- Locally used taxon name
- Where species grows in target region
- Timing of local flowering / fruiting
- Habitat preference
- Topographic preference
- Soil preference
- Geological preference
- Climatic and microclimatic preference
- Breeding system
- Genotypic and phenotypic variation
- Biotic interactions (pests, pathogens, herbivores)
- Archaeological information (e.g. palynology)
- Ethnobotanical information (e.g. traditional uses)
- Conservation status (e.g. Red Data Book status).

Survey of Geographical, Ecological and Taxonomic Data

- Area-specific data
  - geographic, climatic
- Sources of data
  - printed media
  - microfiche
  - on-line databases
  - CD-ROM
  - Internet



Area-Specific Data

- Geographic and climatic data e.g.topography, geology, soil, climate and land use

These can be used to characterize the target area more fully.

Sources of Data

- *Printed Media* - e.g. monographs, revisions, floras, gazetteers, papers, atlases, maps (soil, vegetation, relief, geological, land use, climatic maps etc. - increasingly available in a digital form)

- *Microfiche* - abstracts of papers

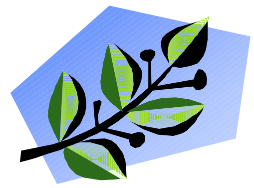
- *On-line Databases* - e.g. CAB abstracts, AGRIS, AGRICOLA, Science Citation Index

- *CD-ROM* - e.g. PlantGene CD, TreeCD, CAB abstracts, Science Citation Index

- *Internet Resources* - numerous resources available for botanists on World Wide Web and Gopher sites. A comprehensive listing is given at the address: http://www.helsinki.fi/kmus/botmenu.html.

Collection of Specimen-Specific Data

– Herbaria and genebanks contain millions of accessions, many in duplicate
– Resources and time are limiting
– Not all data are relevant



Choose carefully which specimens to include
Confirm identification of each specimen

 Ecogeographic Surveys - 24 -

The scope of all ecogeographic investigations will be limited by the availability of time, facilities and materials. The herbaria of the world contain millions of specimens and the number of specimens of any one target taxon can be vast. Since it is not possible to include all specimens of a target taxon in a survey, representative specimens of the target taxon must be selected. Each of these specimens will require identification and a proportion will be selected according to the integrity of the passport data. During the course of a study, Maxted (1995) found that data from about a third of the specimens seen during the study were finally included in the ecogeographic database.

The different stages can be summarised as follows:

- Select representative specimens
- Confirm identification of each specimen
- Check integrity of passport data for each specimen
- Record passport data in ecogeographic database.

Which Specimens to Select?

- Select accessions with detailed passport data
  - Good provenance data are essential
- Select accessions of taxonomic, geographical or ecological interest
  - Positively select accessions to reflect the range of ecogeographic conditions
  - Avoid duplicates

 Ecogeographic Surveys - 25 -

Specimens should be positively selected to represent the breadth of geographic and ecological conditions under which the target taxon is found.

Specimens are more likely to be selected if:

- they have detailed ecogeographic passport data, especially provenance (e.g. longitude/latitude) data.
- they show features of particular taxonomic, ecological or geographic interest, that is, they are odd or rare forms, come from unusual environments or are found on the edges of their natural distributional range.

It is important that emphasis be placed on obtaining reliable specimen provenance data for those specimens to be included in the database. Ideally, only specimens that have either latitude and longitude data available (or for which these data can be established) should be selected for inclusion in the database. In practice, it may be advisable to include specimens with two levels of detail, those for which full latitude and longitude details can be obtained and those with major country subunit detail (i.e. province or state) (Rhoades and Thompson 1992). Specimens that lack even this level of geographical data should not be included unless they are particularly noteworthy taxonomically or in some other way.

Specimen-Specific Descriptors

- Herbarium, etc.
- Collector's name and number
- **Collecting date**
- **Sample identification**
- **Locality**
- **Altitude**
- **Habitat**
- Phenological data
- Soil type
- Vegetation type
- Site slope and aspect
- Land use
- Phenotypic variation
- Pests and pathogens
- Competitive ability
- Palatability
- Vernacular names
- Plant uses

 Ecogeographic Surveys - 26 -

Listed below are the specimen-specific descriptors (data items) that can be obtained from passport data associated with herbarium specimens and germplasm accessions. This list is quite extensive and it is unlikely that all will be recorded during a single survey. Those descriptors that are considered essential for yielding predictive results are shown in **bold.**

- Herbarium, genebank or botanical garden where specimen is deposited
- Collector's name and number
- **Collecting date (to derive flowering and fruiting times)**
- **Sample identification**
- **Locality** (latitude and longitude or even greater detail if possible)
- **Altitude**
- **Habitat**
- Phenological data (presence of flowers and/or fruit)
- Soil type
- Vegetation type
- Site slope and aspect
- Land use and/or farming system
- Phenotypic variation
- Evidence of pests and pathogens
- Competitive ability
- Palatability
- Vernacular names
- Plant uses.

Use of Data Standards

- Data must be recorded
  - consistently
  - using data standards
- Essential for meaningful data analysis and exchange

 VERSION 1.3 ***Ecogeographic Surveys - 27 -***

It is essential to record the data in a consistent way and to use accepted data standards, otherwise, meaningful data analysis and exchange **cannot** be carried out. The use of data standards will also facilitate data recording itself.

The use of data standards is widespread and many standards are relevant in the recording of ecogeographic data, from species and cultivated plant nomenclature to provenance, geological, climatic and biotic data. Therefore, existing standards should be consulted before data recording. A number of organisations and networks are active in setting and promoting data standards, relevant organisations including:

- **International Working Group on Taxonomic Databases (TDWG)**
TDWG is a working group of the International Union of Biological Sciences, set up to develop standards for botanical data to facilitate data exchange. Published standards include the “Transfer Format for Botanic Garden Records” (which includes accession and conservation descriptors) and the “World Geographic Scheme for Recording Plant Distributions”. Under development are standards for habitat, soil, landscape, life form and plant occurrence and status.

- **World Conservation Union (IUCN)** - standards for conservation status of species.

- **International Plant Genetic Resources Institute (IPGRI)** - publishes plant germplasm descriptor lists which include standards for passport, characterization and evaluation data.

Note: data standardisation is covered in more detail in unit 4.

Problems with Data Collection

- Older specimens
  - poorer quality passport data
  - illegible hand-written labels, unfamiliar language used
  - difficulties in determining provenance data because place names may have changed
  - need to infer data (e.g. altitude)

Inferred data should be flagged in the database

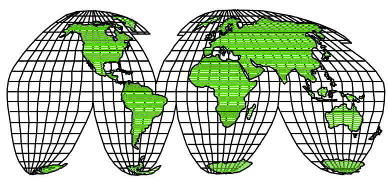
Early plant collectors could not have predicted the detailed analysis that would subsequently be based on the provenance information they recorded. These data are therefore often very basic: site location data may be ambiguous and ecological details missing. Older specimen labels are handwritten, often difficult to read and may be written in an unfamiliar language. Botanists from the herbarium can help here.

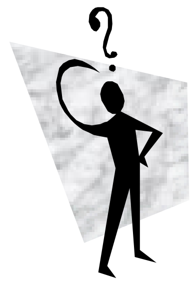
Inference of data

It may be necessary to infer various features (latitude and longitude, geology, soil, altitude, etc.) of collection sites from location data by reference to appropriate maps. Whether this is possible will depend on the precision of the location data available, the topography of the collection area and the precision of the environmental data required. For example, if the collecting site is situated in a mountainous area, then the altitude is likely to vary markedly within short distances and so estimates of site altitude will be misleading unless the location is precisely specified. A gazetteer of local geographical names and locations is helpful in inferring a specimen's provenance. There is no comprehensive world gazetteer yet available, but other publications contain extensive gazetteers such as the Times Atlas of the World (Anon. 1988). Older maps, atlases, gazetteers and even travel books can be a useful source of localities where the original name or boundary has changed. In some cases it may even be necessary to infer the provenance data from the collector's original notebooks. In all cases, data which have been inferred should be flagged in the database to distinguish them from other types of data.

Problems with Data Collection

- Older specimens do not necessarily reflect the current distribution of the species





Another problem with using older specimens extensively is that, while providing accurate information about historical distributions, misleading inferences about contemporary populations can be drawn. This might present a special problem when trying to locate populations of threatened, therefore rare and restricted taxa. For example, van Slageren (1990) indicated that populations of *Aegilops uniaristata* seem unlikely to have survived in Turkey beyond 1900, however old herbarium specimens of this taxon from this country are numerous and widespread.

In general, it is much easier to record curatorial or geographic data than ecological data from herbarium specimens. This may restrict the ecological conclusions that can be drawn from the data set.

Data Verification

- Assess completeness of the data set
  - certain analyses are not possible if it is incomplete
- Check for errors
  - duplications
  - typing errors
  - inconsistencies e.g. latitude/longitude

Errors can be avoided with careful system design

Assess completeness of data set

After data entry is complete, an assessment should be made of the completeness of the data set as there are commonly missing data in many fields. If many of the records have significant amounts of missing data, certain kinds of mathematical analysis will not be possible or will give misleading results. To help determine the completeness of the data set, a record should be flagged when all possible information has been collected.

Check for errors

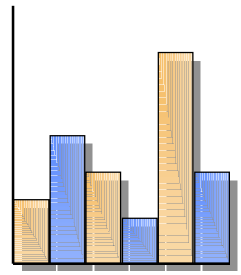
It is important to check for errors that may have been entered in the database. A lot of errors can be avoided through the thoughtful design of the system, *e.g.* data validation on data entry, storing default values, lower case / upper case conversion, range checking in numeric data, etc.

A useful way of identifying typing errors is to sort (or index) the records on a particular field and then examine the data for that field. Errors are then much more obvious, e.g. *Gossyppium* instead of *Gossypium*. The process can be repeated on other fields or a combination of fields. Duplicate specimens should be located by sorting on the relevant fields (e.g. accession number, collector’s number, other numbers associated with the accession). This is advisable as many collectors send duplicate sets of herbarium specimens to different international herbaria and germplasm accessions are commonly duplicated.

Errors in mapping latitude and longitude data can be detected if particular localities are shown up as obvious outliers in unlikely places.

Data Analysis

- Tables and bar charts
- Multivariate statistical analysis
- Maps
  - – Enclosed line maps
  - – Dot distribution maps
  - – Contour maps
- Geographic Information Systems (GIS)



The ecogeographic database can be analysed to help identify the geographical locations and habitats favoured by the target taxon. The types of analysis that can be performed range from the production of simple bar charts to complex statistical analysis and the use of Geographic Information Systems. The types of analysis can be summarised as follows:

- **Tables and bar charts**
- **Multivariate statistical analysis**
- **Maps**
  - Enclosed line maps
  - Dot distribution maps
  - Contour maps
- **Geographic Information Systems (GIS)**

Tables and Bar Charts

IPGRI

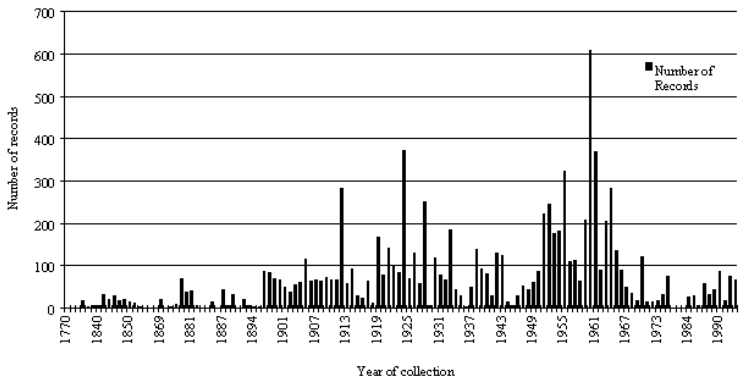
- Used to display relative frequencies
  - of specimens from different sites
  - of characters from different sites
  - of specimens/characters against environmental variables
- Can identify ecological niche of target taxon
- Can show ecotypic adaptation in wild and cultivated species

One of the simplest forms of ecogeographic data analysis is to calculate the number of specimens collected from geographically or ecologically distinct sites (e.g. climate, soil types, aspect, shading characteristics, habitat, etc.). This can be done easily with database or spreadsheet software. The results can be expressed on tables, bar charts, pie-charts, graphs, etc. Data arranged in this fashion can help identify the particular niche that the target taxon prefers.

Correlation of the abundance of a taxon with ranges of particular environmental descriptors (e.g. altitude, latitude and soil pH) can be used predictively. Correlation of morphological characters with ranges of particular environmental descriptors can help indicate possible ecotypic adaptation, both in wild and cultivated material.

Pie-charts can be used to display the relative frequency of a character in different places. They are commonly used to show allelic frequencies in populations at different geographical locations, but have also been used to display morphological variation, e.g. stigma colour in *Crocus scepusiensis* (Rafinski 1979). New (1958) used pie-charts to demonstrate seed coat variation in *Spergula arvensis*, the relative size of the circles indicating the size of the population sampled; different sized circles could also be used to indicate the number of database records from which the frequency has been estimated. Pie-charts can also be used to compare the distribution of specimens with physical characteristics, e.g. altitude, temperature or soil type. Daday (cited by Jones, 1973) used pie-charts to show the relationship between the distribution of cyanogenic forms of *Lotus corniculatus* in Europe and January isotherms.

Histogram -
Collecting Frequency



This histogram shows the number of herbarium specimens of Papilionoid legumes collected in Madagascar between 1770 and 1993. These are held at the Herbier, Laboratoire de Phanérogamie, Muséum National d'Histoire Naturelle, Paris (Sawkins 1995).

Multivariate Statistical Analysis

- Analyses more than one variable at a time
- Many methods available
  - choose the most appropriate
  - only works with robust data
- Hierarchical clustering methods
  - divide objects into groups based on similarity
  - defined groups can be used in other analysis

In many cases, more complex statistical analyses are needed which deal with more than one variable at a time - these are called multivariate analyses. For instance, it may be necessary to divide the collecting sites into separate groups according to the overall climate (e.g. mean annual and monthly rainfall, mean minimum/maximum temperature, etc.), rather than use a single environmental variable such as rainfall.

A wide range of multivariate analysis techniques exist for analysing groups of ecogeographic or taxonomic variables. No single method will give predictive results in all situations - the actual method selected will depend on the aim of the analysis, the structure of the data, the availability of software and any advice given by a statistics expert. These analyses are only possible with robust data.

There are two main types of multivariate analysis of particular relevance to ecogeographic data - *hierarchical clustering* methods and *ordination* methods.

Hierarchical clustering

This divides objects into different groups, called *clusters*, on the basis of their similarity; two objects in the same cluster are more similar than two objects in different clusters. The objects may be plant populations, collecting sites or other type of sites. The clusters identified can be used as entities in further analysis, e.g. distribution maps and regression analysis.

The clustering process may be of two types: (1) *agglomerative* - in which objects are joined into progressively larger clusters; or (2) *divisive* - in which the cluster of all objects is progressively divided into smaller clusters (not discussed here).

Hierarchical Clustering (cont^{d})

- **Stages in clustering**
  - Find the two objects which are most similar
    - choose most appropriate similarity coefficient
  - Place in a cluster and treat as an object
  - Repeat until all objects are in one cluster
- **Composition of clusters depends on**
  - methods used to determine similarity between clusters
  - other rules for membership of a cluster

The stages of agglomerative clustering are as follows with examples of the statistical methods used:

- Find the two objects which are most similar
  - for nominal scale data - simple matching coefficient
  - for continuous scale data - Euclidean distance
  - for binary scale data - Jaccard coefficient
  - for nominal and continuous scale data - Gower’s general similarity coefficient
- Place the two most similar objects in a cluster - treat as an object in the subsequent analysis
- Repeat the analysis to identify the next most similar objects/clusters and place these in a cluster as before
- Repeat the process until only one cluster remains.

There are a number of different methods for calculating the similarity between clusters (as opposed to between individuals) and this can affect the clustering. One way to produce geographically compact clusters is to disallow membership if an object/cluster falls outside a pre-determined distance.

It should be emphasised that the usefulness of any defined clusters depends on (1) the descriptors used; (2) the quality of the data; (3) the statistical methods used to define the clusters; (4) the subsequent use of the clusters in other analyses.

Ordination Methods

- Principal components analysis (PCA)
  - produces new quantitative variables which reflect the variation in the original descriptors
- Principal coordinates analysis (PCO)
  - similar to PCA; for quantitative and qualitative data
- Canonical variate analysis (CVA)
  - distinctiveness of pre-defined groups

The second type of multivariate analysis possible with ecogeographic data is *ordination* or *multidimensional* scaling. These are analyses which do not produce groups of objects as with clustering methods but instead bring out the overall similarity using geometric methods.

The best known ordination method is Principal Components Analysis (PCA) which can be used for quantitative data. Using this method, a long list of descriptors can be reduced to a few, entirely new, independent and linear variables (synoptic variables or principal component axes) which contain most of the variability from the original descriptors and which can be plotted just like any other continuous descriptor. So, for example, plant populations which have similar principal component scores will tend to be similar with regard to the descriptors contributing to each principal component score.

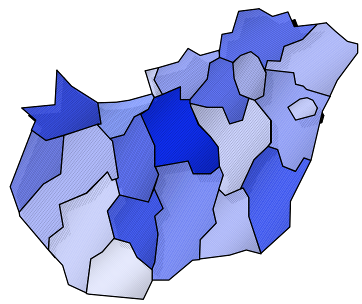
For qualitative and quantitative data, Principal Coordinates Analysis (PCO) can be used. This uses a similarity matrix of the type generated in cluster analysis to perform the geometric plots which are then analysed in the same way as for PCA.

Canonical variate analysis (CVA) is another ordination technique which produces canonical variates (the equivalent of principal components) which are a measure of the distinctiveness of a pre-defined group compared with other groups in the analysis. In computing the distinctiveness, CVA takes into account the variation within a group. An example could be a set of landraces, each group collected from a different country.

Other widely used ordination techniques include non-metric multidimensional scaling and detrended correspondence analysis.

Maps

- Enclosed line maps
- Dot distribution maps
- Contour maps



 Ecogeographic Surveys - 37 -

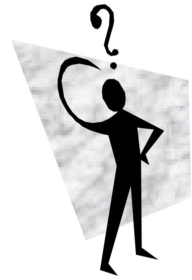
A powerful way to identify the geographical locations and habitats favoured by the target taxon is to produce distribution maps showing the location of collecting sites and the data obtained, including data derived from statistical analysis. These distribution maps can be superimposed on topographical, climate, geological or soil maps.

There are three approaches that can be taken:

- **Enclosed line maps** - shading or enclosing an area with a single line to show the distribution of the target taxa
- **Dot distribution maps -** presence is indicated using dots on the map
- **Contour maps** - e.g. isoflor map: distribution of plant taxa indicated by joining together points which have the same number of taxa.

Enclosed Line Maps

- Simple approach but can give misleading results
  - no indication of frequency
  - no local variation
  - problems with isolated occurrences



In enclosed line maps, an area is shaded or enclosed with a single line to show the presence of the taxon. This is a simple approach but the results can be misleading because:

- no indication is given of the frequency of the target taxon within the enclosed area
- any local variation due to ecological/geomorphological factors cannot be shown
- an isolated occurrence of the taxon might wrongly suggest a distribution throughout the region.

Dot Distribution Maps

- Preferred choice for showing distribution patterns
- Can superimpose additional information
- Mapping software very useful for flexibility in the display of data

 VERSION 1.3 ***Ecogeographic Surveys - 39 -***

When distribution patterns require deatiled examination, dot distribution maps are generally preferred to line maps. With a dot distribution map, each point represents the actual locality where the sample was collected. Additional morphological or ecological information about the sample or the collecting site can be superimposed onto the dots in the form of different symbols to show combinations or frequencies of the character states. These symbols may take the form of circles, bars, pie-charts, bars, bar-graphs, etc., depending on the data presented. For instance, pie-charts can be used to display the relative frequency of a character in different places and differently shaded circles can be used to display different soil types.

Displaying the distributions of taxa can be helped enormously by the use of mapping software. Mapping programs will allow the import of latitude and longitude coordinates from the ecogeographic database and their plotting onto computerised maps, the scale of which can be customised to display the distribution to best effect.

Dot Distribution Map

 Ecogeographic Surveys - 40 -

This map shows the distribution of *Vicia serratifolia* throughout Europe, Southwest Asia and North Africa (Maxted 1995).

Application of Dot Distribution Maps

- Location of germplasm samples and herbarium records
- Passport, characterisation and evaluation descriptors
- Membership of different species, cultivars or landraces
- Synoptic descriptors
- Within-population variation

 VERSION 1.3 ***Ecogeographic Surveys - 41 -***

- **Location of germplasm specimens and herbarium (literature) records**

These can be plotted on the same map using different symbols. Areas clearly stand out where a species has been recorded but for which there are no germplasm samples available.

- **Passport, characterisation and evaluation descriptors**

Certain descriptors can be plotted using different symbols such as circles or lines. Qualitative descriptor states can be represented by different shadings in circles, quantitative descriptors by different size circles or line length. Examples: soil type, altitude, population size, threat of genetic erosion, plant height, leaf shape, biochemical markers, etc. This can show the ecological preferences of a group or ecotypic adaptation.

- **Membership of different species, cultivars or landraces**

Using different symbols to represent the different groups, it is possible to demonstrate areas of high taxonomic diversity, co-occurrence of species, species rarity and endemism.

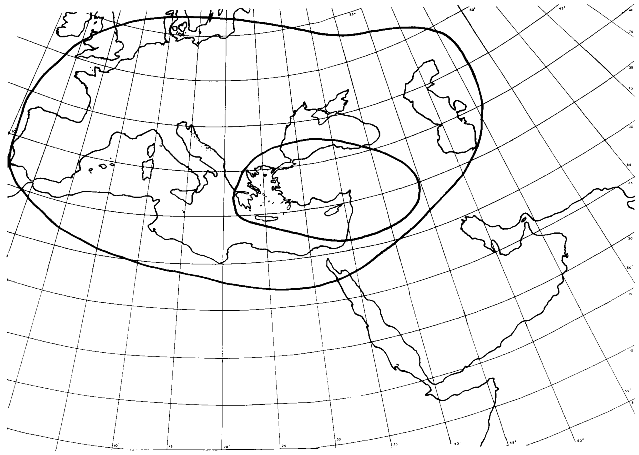
- **Synoptic descriptors**

Synoptic descriptors from the multivariate analysis can be plotted like any other quantitative descriptor and displayed with other data to show possible correlation in their distribution

- **Within-population variation**

Statistical measures of variation can be co-plotted with the quantitative descriptor state, for example, mean leaf length plotted with the standard deviation. For qualitative descriptors (e.g. fruit colour), a pie-chart can be be displayed show the frequency of each descriptor state.

Contour Map



Contour maps showing the distribution of plant taxa can be produced by joining together points which have the same number of taxa. Such maps are sometimes called *isoflor* maps. Isoflor maps can also be produced for the distribution of subspecies, landraces. etc.

Contour maps can be constructed for other types of ecogeographic data, including morphological characters (*isophene* maps) and other environmental characters (e.g. temperature - *isotherm* maps).

The advantage of contour maps is that they can be compared directly with other contour maps and the associations between the different ecogeographic characters are easier to visualise.

The isoflor map above shows the distribution of *Vicia* sect. *Wigersia* which comprises two species, *V. cuspidata* and *V. lathyroides. V. cuspidata* is restricted to Southeast Europe and Southwest Asia and *V. lathyroides* extends to North Africa and Western Europe (Maxted 1995).

Geographic Information Systems (GIS)

- Dedicated to handling spatial data and logically related, non-spatial data
- Highly adaptable and versatile
- Powerful analysis tool

 Ecogeographic Surveys - 43 -

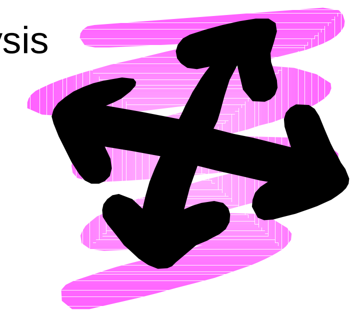
The analysis of ecogeographic data can be much easier when a Geographic Information System (GIS) is used. A GIS is a database management system dedicated to handling geographically referenced spatial data (such as coordinates for site location, topography) in a graphical form together with logically related, non-spatial data (e.g. species name, morphological characters).

A GIS is a highly adaptable mapping system, which can easily cope with a wide range of geographical, ecological and biological data sets. Once cartographic data from existing maps (which will often be at different scales), aerial photographs, field surveys, remote sensing, etc. are in digital form, they can be manipulated and analysed in various ways. For example, different data sets can be superimposed on a computer screen to enable direct visual comparisons.

Among other uses, a GIS makes it easier to infer features of collecting sites for which no data were recorded at the time of collection, and to locate areas with particular combinations of ecological characteristics which could be targeted for future collecting. Computer programmes are already available which will carry out such analyses.

GIS Capabilities

- Geometric correction
- Digital terrain model analysis
- Interpolation
- Overlay analysis
- Proximity analysis
- Computation of statistics
- Location



- **Geometric correction**. The scale and projection of different maps can be changed to make them comparable and amenable to overlay analysis.

- **Digital terrain model analysis**. The altitude contours on a topographical map may be used to produce maps of slope, aspect, shaded relief, etc. These may constitute important additional provenance information.

- **Interpolation.** Point data may be used to create contour maps, such as isoflor maps or maps showing areas where morphologically similar specimens have been collected.

- **Overlay analysis**. Different maps of the same area may be combined to produce a new map. For example, maps of slope, soil, wind speed and vegetation cover may be overlaid to synthesize a map of potential soil erosion or desertification, which may be useful in identifying areas at particular risk of genetic erosion. If some passport data are missing (e.g. the soil type or altitude at the collecting point), they may be derived by overlaying the species distribution map on different thematic base maps.

- **Proximity analysis**. Buffers may be generated around features such as wells, villages and roads to determine the accessibility of potential reserves.

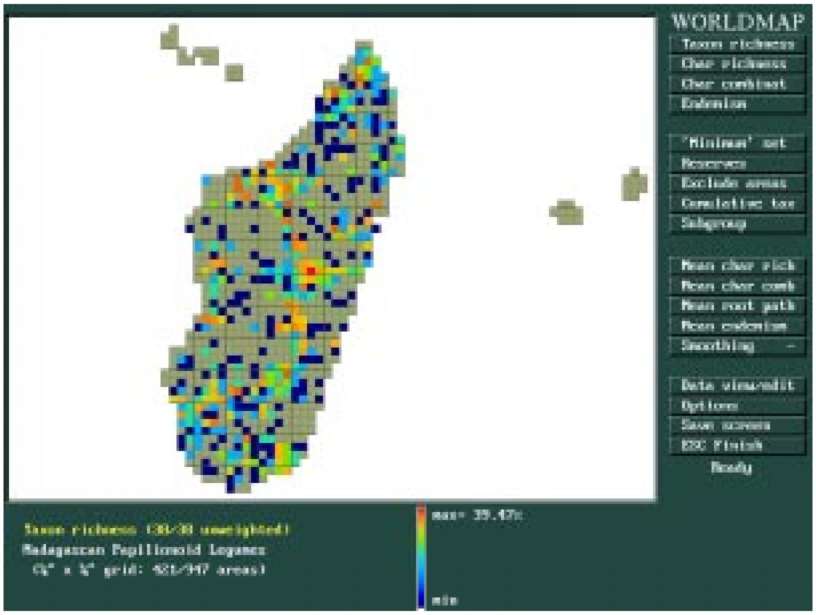
- **Computation of statistics.** Means, counts, lengths and areas may be calculated for different features e.g. the percentage of the total occupied by different vegetation or soil types.

- **Location.** Entities having defined sets of attributes may be located (e.g. all specimens recorded from particular soil types).

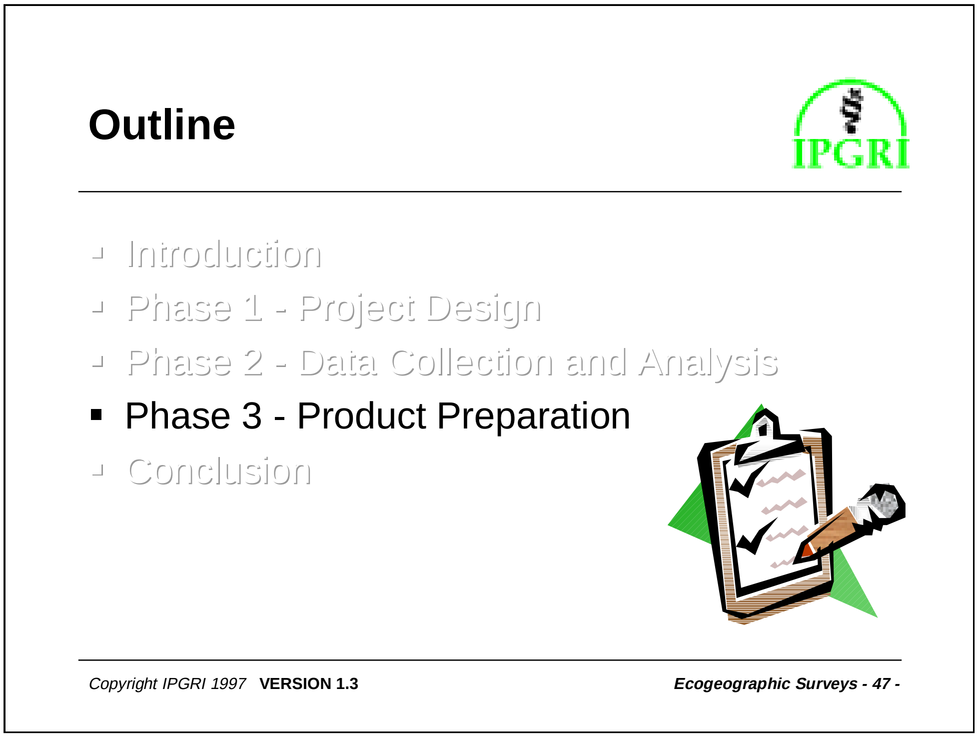
Example - WORLDMAP

- Graphical, analytical tool
  - Used to identify priority areas for conservation
- Displays distribution data with cladistic information
- Priority areas identified according to one or more criteria

WORLDMAP (Williams and Humphries 1994) is an example of a graphical, analytical software tool that can be used for identifying priority areas for conservation. It allows the input of distributional data together with cladistic information so that priority areas can be identified according to one or more criteria. Such priority areas are displayed graphically in the form of a map.



In the above example, distributional data are displayed for herbarium specimens of Papilionoid legumes collected in Madagascar between 1770 and 1993 (Sawkins 1995). The criterion applied here is taxon richness, that is, the sum of the number of *taxa* in an area. Using this criterion, the majority of the diversity is shown to be in the central plateau region with smaller areas in the north, northwest, southeast and southwest of the island. The areas in which the highest diversity occurs are open savanna and seasonally dry deciduous forests.



Phase 3 - Product Preparation

Data Synthesis

The data collected and the analyses performed throughout the survey are brought together in the ecogeographic database, conspectus and report. A good appreciation of the completeness of the data set is stressed.

Ecogeographic Conspectus

The contents of an ecogeographic conspectus are discussed and a checklist of information to include is given.

Ecogeographic Report

The contents of an ecogeographic report are summarised together with a checklist of recommended discussion points to be included. A list of criteria used in the setting of conservation priorities is given.

Data Synthesis

- Final synthesis of all data and analyses
- Products
  - ecogeographic database
  - ecogeographic conspectus
  - ecogeographic report

Important to know the completeness of the data set

As mentioned previously, there are three basic outputs of an ecogeographic survey or study:

- **The ecogeographic database** - containing the raw data for each taxon
- **The ecogeographic conspectus** - summarising the data for each taxon
- **The ecogeographic report** - discussing the contents of the database and proposing future conservation strategies and collecting missions as appropriate.

The final production phase of the project commences with the synthesis of all the data collected during the study to produce the final conspectus and report.

The researcher should know how complete the database is and how complete the collections surveyed are in terms of covering the target area.

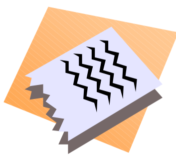
A particular problem is as follows. If a particular habitat is under-represented in the database, is it because:

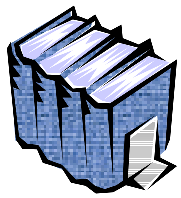
- the taxon is absent from that habitat?
- that type of habitat has not been sampled or sampled inadequately?
- the target taxon has not been recognised in such a habitat?

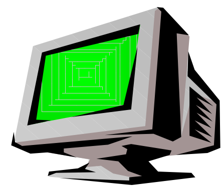
This problem must be considered; otherwise, the results of the analysis and the inferences drawn from them could be misleading.

Ecogeographic Conspectus

- Summarises all the available information collected in the survey
- Arranged in order of plant name
- Provides summaries for each geographical subunit







Ecogeographic Surveys - 49 -

The conspectus is a summary of all the available ecological, geographical and taxonomic information collected in the survey for the target taxon through part or all of its range. The conspectus is arranged in order of plant names (listed either alphabetically or systematically) and, where appropriate, contains an index to the taxa included in the study.

If the scope of the investigation is broad in the geographical or ecological sense, it may be necessary to provide a summary of the ecogeographic data for each geographical or ecological subunit. For instance, in Edmonds' survey of African Corchorus (Edmonds 1990), the flowering time for each species in each country was listed rather than providing one time range for the whole continent. This increased the predictive value of the survey

Conspectus - Information to Include

- Accepted taxon name, author(s), etc.
- Reference to published descriptions
- Morphological descriptions or identification keys
- Phenology, flowering season
- Ethnobotanical notes
- Geographical distribution
- Distribution maps
- Ecological, geographical and taxonomic notes
- Conservation notes

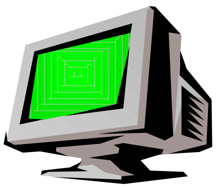
 Ecogeographic Surveys - 50 -

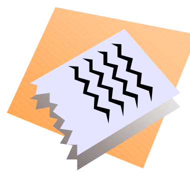
Where possible, the following information is included in the conspectus:

- Accepted taxon name, author(s), date of publication, place of publication
- Reference to published descriptions and iconography
- Short morphological descriptions or keys for important taxa or those that may be difficult to identify
- Phenology, flowering season
- Ethnobotanical notes, e.g. vernacular names and local uses
- Geographical distribution, i.e. countries from which the taxon is recorded, including reliable records from the literature
- Distribution maps (preferably dot distribution)
- Ecological notes, including: altitude (minimum and maximum), habitat, topographic, soil, geological, climate and microclimate preference, biotic interactions and other habitat details
- Geographical notes, including an interpretation of the taxon's geographical distribution
- Taxonomic notes, including notes on any distinct genotypic and phenotypic variation within the taxon
- Conservation notes, containing an assessment of the variability currently conserved, the potential genetic erosion faced and the conservation status of the taxon in the field.

Ecogeographic Report

- The ecogeographic report contains
  - a discussion of the database and conspectus
  - conclusions regarding the group's ecogeography
  - a concise list of conservation priorities





The report discusses the contents of the database and conspectus and draws general conclusions concerning the group's ecogeography, presenting a concise list of conservation priorities.

Report - Discussion Points

- Taxon delimitation, classification
- Taxon ecology and uses
- Database structure; hardware/software used
- Taxon conservation status
- Suggested conservation priorities and strategy

 VERSION 1.3 ***Ecogeographic Surveys - 52 -***

If possible, the following points should be covered:

- The delimitation of the target taxon
- The classification of the target taxon that has been used, and why
- The mode of selection of representative specimens
- Target taxon ecology
- Target taxon phytogeography, discussion of the distribution patterns and a summary of the distribution in tabular form
- Any interesting taxonomic variants encountered during the study
- The current and potential uses of the target taxon
- The relationship between the cultivated species and their wild relatives
- Any particular identification problems associated with the group, presentation of identification aids to vegetative, floral and fruiting specimens
- The choice of hardware and software
- The ecogeographic database file structures and inter-relationships
- Database content
- *In situ* and *ex situ* conservation activities associated with the target taxon, including the extent of diversity already conserved
- Any threat of genetic erosion facing the group
- Priorities and suggested strategy for future conservation of the target taxon (see next slide).

The ecogeographic conspectus may be included within the report as an appendix or as a separate entity.

Criteria Used in Setting Conservation Priorities

IPGRI

- Biological
  - Richness
  - Rarity
  - Threat
  - Representativeness
  - Function
- Social/Political
  - Utility
    - economic, scientific, social, cultural and religious use
  - Feasibility
    - political, economic, logistical and institutional considerations

 Ecogeographic Surveys - 53 -

Not all populations, species or ecosystems identified in the analysis will receive priority status for conservation. This depends on the criteria selected and how they are applied. These can be divided into two groups: biological and social / political (Miller *et al.* 1995).

- **Biological**
  - *Richness* - the number of populations / species / ecosystems in the area
  - *Rarity*
  - *Threat* - the degree of danger
  - *Representativeness* - how closely an area represents a defined ecosystem
  - *Function* - the degree to which a population / species / ecosystem affects others to exist
- **Social/Political**
  - *Utility* - economic, scientific, social, cultural and religious use
  - *Feasibility* - political, economic, logistical and institutional considerations

Priority-setting is a complex process, from the method used in the analysis of a geographic area to the selection and application of the different criteria. However, it is recognised that feasibility - i.e. the likelihood of success of a conservation objective in the given social and political climate - is a key factor in priority-setting and the allocation of resources.

Outline

- Introduction
- Phase 1 - Project Design
- Phase 2 - Data Collection and Analysis
- Phase 3 - Product Preparation
- Conclusion

Herbaria, genebanks and botanical gardens are storehouses of botanical data as much as they are of plants - pressed, frozen or live. These data can be used to facilitate and order plant conservation. Analysis of a taxon’s geography, ecology and taxonomy is a necessary prerequisite for assessing its conservation status and permits the prediction of which areas and habitats the taxon is likely to be found in. Once located, populations of the taxon can be monitored, sampled if necessary and effectively conserved. Ecogeographic studies and surveys will always be limited by time and resources and it will be impossible to collate every piece of information on a taxon’s geography, ecology and taxonomy. However, if they are carefully planned and executed, the data collated are predictive and help the conservationist to identify potential collection and conservation priorities more effectively.