Refactoring DNA Surveillance
Species identification with DNA
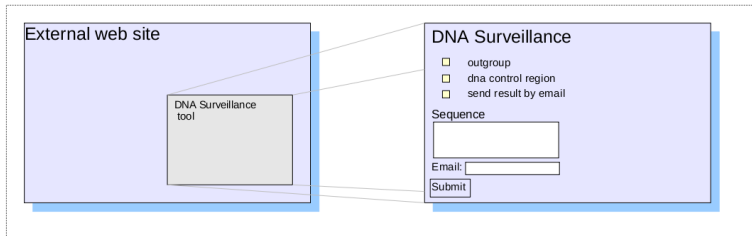
Waltraud Sichart
Supervisor: Howard Ross

December 10, 2009

Overview

Introduction The Java TreeBuilder application The database behind the web site The DNA Surveillance (Admin) web site
●○ ○○ ○○ ○○○

What is DNA Surveillance

DNA Surveillance should be a tool, that can be inserted into other web sites (taxon specialist groups in particular),
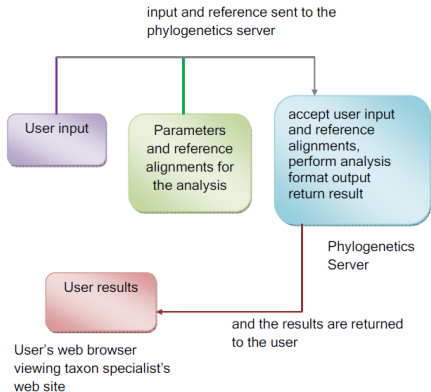

where users submit sequences and get a tree with different species as result.

The look and feel of the whole page is under the control of the taxon specialist or their assistant.

Introduction The Java TreeBuilder application The database behind the web site The DNA Surveillance (Admin) web site

How should that work

When a user performs an identification

input and reference sent to the phylogenetics server

User input

Parameters and reference alignments for the analysis

accept user input and reference alignments, perform analysis format output return result

Phylogenetics Server

User results

and the results are returned to the user

User's web browser viewing taxon specialist's web site

Important constituent parts

Web interfaces

- For the users to submit sequences
- For the administrators to add reference alignments to the database

Storage

- Database for the reference alignments and other information

Analysis

- Application to perform analysis
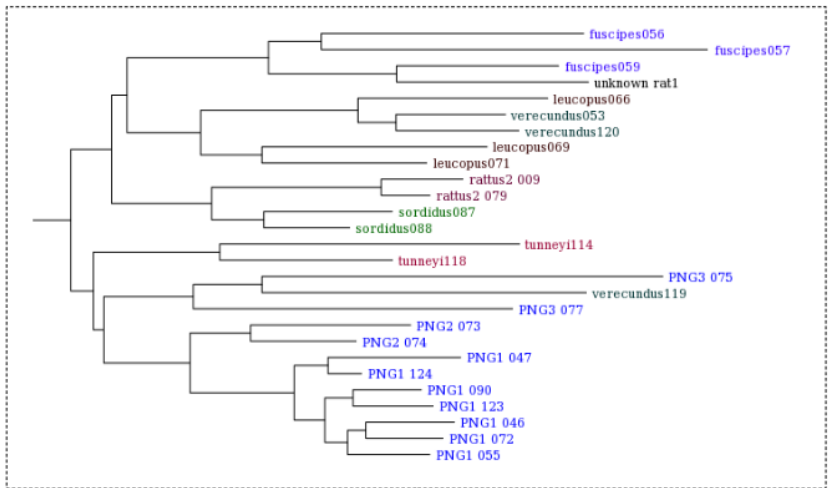- Phylogenetics Server to run the application

Program sequence for analysis

▷ Build profile alignment with ClustalW

1. Set information for the profile alignment
2. Create command as String array and execute it
3. Wait for process to finish

Program sequence for analysis

▷ Build profile alignment with ClustalW

1. Set information for the profile alignment
2. Create command as String array and execute it
3. Wait for process to finish

▷ Create distance matrix and neighbour joining tree

1. Read in the new alignment file and choose a distance model
2. Build a neighbour joining tree and root it
3. Add tip colours as attributes to the tree
4. Show tree and save it as .tre file

Introduction   The Java TreeBuilder application   The database behind the web site   The DNA Surveillance (Admin) web site
○○   ○●   ○○   ○○○

The Result Tree

Introduction   The Java TreeBuilder application   **The database behind the web site**   The DNA Surveillance (Admin) web site
○○   ○○   ●○   ○○○

The MySQL database

Reasons of setting up a MySQL database

+ Setting up the database with phpMyAdmin
    ▷ Very fast and simple to set up
    ▷ Easy to maintain
        → Even without/less MySQL knowledge

Introduction The Java TreeBuilder application **The database behind the web site** The DNA Surveillance (Admin) web site
○○   ○○   ●○   ○○○
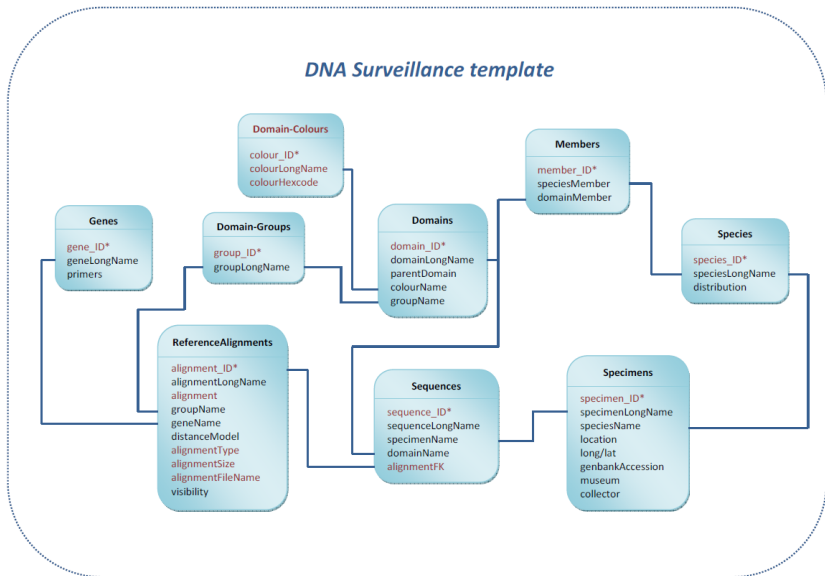
The MySQL database

Reasons of setting up a MySQL database

+ Setting up the database with phpMyAdmin
  ▷ Very fast and simple to set up
  ▷ Easy to maintain
    → Even without/less MySQL knowledge

InnoDB as table storage engine

+ Foreign keys and relationship contrains

− Different relationships between tables, so it takes more time to design the data models

Introduction   The Java TreeBuilder application   **The database behind the web site**   The DNA Surveillance (Admin) web site
○○   ○○   ○●   ○○○

The database model

DNA Surveillance template

The web site is the interface by which administrators interact with the database

It should be ...

- easy to use and everything should be good readable
- flexible so it can be customized to fit the user's individual needs
- usable for all databases that were created of a specific template
- secure to protect the stored information

This aims were achieved with various design points ...

Colours and text

- Colours with high contrast for readable text so that the page is easy to use
- Mostly used font size is 14pt to assure good readability for serif fonts and sans serif fonts

Introduction   The Java TreeBuilder application   The database behind the web site   **The DNA Surveillance (Admin) web site**
○○   ○○   ○○   ○●○

Design Points

Colours and text

- Colours with high contrast for readable text so that the page is easy to use
- Mostly used font size is 14pt to assure good readability for serif fonts and sans serif fonts

Layout

- Clearly arranged layout to facilitate the usage
- Dynamic table layout to fit various numbers of columns and rows
- Diffent row colours for odd and even rows to increase readability

Introduction   The Java TreeBuilder application   The database behind the web site   **The DNA Surveillance (Admin) web site**
○○   ○○   ○○   ○○●

Design Points

Flexibility

- All layouts were created using cascading style sheets
- All Elements have their width defined in percentages to match all screen resolutions
- The to-use database name is recorded in the user database and is loaded after the user validation

Introduction   The Java TreeBuilder application   The database behind the web site   **The DNA Surveillance (Admin) web site**
○○   ○○   ○○   ○○●

Design Points

Flexibility

- All layouts were created using cascading style sheets

- All Elements have their width defined in percentages to match all screen resolutions

- The to-use database name is recorded in the user database and is loaded after the user validation

$\rightarrow$ The interface can be used for all databases that were created of a specific template

Introduction The Java TreeBuilder application The database behind the web site **The DNA Surveillance (Admin) web site**
○○   ○○   ○○   ○○●

Design Points

Flexibility

- All layouts were created using cascading style sheets
- All Elements have their width defined in percentages to match all screen resolutions
- The to-use database name is recorded in the user database and is loaded after the user validation
- → The interface can be used for all databases that were created of a specific template

Security measures

- Users get their own username, password and a database name
- The password in encrypted using the md5 algorithm with a salt
- Only the encrypted password is stored in the database

▸ Go to web site

Web site

- Import function for alignment details

Introduction The Java TreeBuilder application The database behind the web site The DNA Surveillance (Admin) web site
oo   oo   oo   ooo
Future prospects

Web site

- Import function for alignment details

Java application

- Get information about reference alignments, distance models and tip colours out of database
- Rooting for an outgroup

Introduction **The Java TreeBuilder application** **The database behind the web site** **The DNA Surveillance (Admin) web site**

Acknowledgments

Acknowledgments

Howard Ross
Pui Shan Wong
Bioinformatics Institute

Introduction **The Java TreeBuilder application** **The database behind the web site** **The DNA Surveillance (Admin) web site**
○○ ○○ ○○ ○○○

Acknowledgments

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Please log in to DNA Surveillance Admin

Username: [   ]

Password: [   ]

[Log in]

Introduction The Java TreeBuilder application The database behind the web site The DNA Surveillance (Admin) web site
○○   ○○   ○○   ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Select a database

InspectorFoode ▾

Submit

Introduction   The Java TreeBuilder application   The database behind the web site   The DNA Surveillance (Admin) web site
○○   ○○   ○○   ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Welcome to DNA Surveillance admin
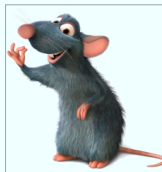
Note: This site uses cascading style sheets, cookies and javascript.

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Current specimens

specimenLongName	speciesName	location	long/lat	genbankAccession	museum	collector
☐ seq1	Bos taurus	Mexico	120W15N	NC_006853	ROM	HAR
☐ seq2	Ovis aries	Mexico	120W15N	NC_001941	ROM	HAR
☐ seq3	Capra hircus	Mexico	120W15N	NC_005044	ROM	HAR
☐ seq4	Equus caballus	Mexico	120W15N	NC_001640	ROM	HAR
☐ seq5	Equus asinus	Mexico	120W15N	NC_001788	ROM	HAR
☐ seq6	Lama glama	Mexico	120W15N	AY535253	ROM	HAR
☐ seq7	Lama pacos	Mexico	120W15N	AJ566364	ROM	HAR
☐ seq8	Camelus bactrianus	Mexico	120W15N	NC_009629	ROM	HAR
☐ seq9	Camelus dromedarius	Mexico	120W15N	NC_009849	ROM	HAR
☐ seq10	Sus scrofa	Mexico	120W15N	NC_000845	ROM	HAR
☐ seq11	Canis familiaris	Mexico	120W15N	NC_002008	ROM	HAR

Options

[Add specimens] [Edit specimens] [Delete specimens]

[Import specimens] [Show table structure]

[Go back to species] [Go to genes] [Finish]

Introduction   The Java TreeBuilder application   The database behind the web site   The DNA Surveillance (Admin) web site
○○   ○○   ○○   ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Field	Type	Value
specimenLongName	varchar(100)	seq1
speciesName	varchar(100)	Bos taurus
location	text	Mexico
long/lat	text	120W15N
genbankAccession	varchar(50)	NC_006853
museum	varchar(50)	ROM
collector	text	HAR

Submit   Cancel

Introduction | The Java TreeBuilder application | The database behind the web site | The DNA Surveillance (Admin) web site
○○ | ○○ | ○○ | ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Import Specimens from file:

All entries in the file must be tab-separated and the last entry in each line must be a semicolon.

The default entries are read in serial in the order:
specimen name **TAB** species name **TAB** location **TAB** long/lat **TAB** genbank accession **TAB** museum **TAB** collector **TAB** ;

If more columns are added to the table please ensure that the type of the import fields is correct, else errors occure (i.e. INT only allows numeric input)

+ Notes ...

Select textfile and click on submit button:

[   ] Browse...

Submit   Cancel

Introduction | The Java TreeBuilder application | The database behind the web site | The DNA Surveillance (Admin) web site
○○ | ○○ | ○○ | ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Structure of Table Specimens

	Field	Type	Null	Action
	specimenLongName	varchar(100)	NO	
	speciesName	varchar(100)	NO	
☐	location	text	NO	✏ ✗
☐	long/lat	text	NO	✏ ✗
☐	genbankAccession	varchar(50)	NO	✏ ✗
☐	museum	varchar(50)	NO	✏ ✗
☐	collector	text	NO	✏ ✗

⤴ Check All / Uncheck All *With selected:* ✗

⊞ Add field ⦿ At End of Table ○ After [speciesName ∨] [Go]

[Go back to table]

> Fields without checkboxes are fixed columns and can not be edited or dropped.
> Each field name must be unique.

Introduction   The Java TreeBuilder application   The database behind the web site   The DNA Surveillance (Admin) web site
○○   ○○   ○○   ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Current reference alignments

alignmentLongName	groupName	geneName	distanceModel	visibility
☐ References	TypesofAnimals	cytochrome B	HKY	public

Options

| Add alignment | Edit alignment | Delete alignment |

| Add alignment details | Edit alignment details | Show alignment details |

| Go back to genes | Finish |

Introduction   The Java TreeBuilder application   The database behind the web site   The DNA Surveillance (Admin) web site
○○   ○○   ○○   ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Control

home
logout
To Database

Layout

look and feel (admin)

Database

domain-groups
domains
species
specimens
genes
reference alignments

Alignment Details

Discription ...

Current alignment for reference alignment: `References`

```
>seq1
ATG···
ACTAACATTCGAAAGTCCCACCCACTAATAAAAATTGTAAACAATGCATTCATCGACCTTCCAGCCCCATCAAACATTTCATCATGA
TGAAATTTCGGTTCCCTCCTGGGAATCTGCCTAATCCTACAAATCCTCACAGGCCTATTCCTAGCAATACACTACACATCCGACAC
AACAACAGCATTCTCCTCTGTTACCCATACTGCCGAGACGTGAACTACGGCTGAATCATCCGATACATACACGCAAACGGAGCT
TCAATGTTTTTTATCTGCTTATATATGCACGTAGGACGAGGCTTATATTACCGGTCTTACACTTTTCTAGAAACATGAAATATTGGA
GTAATCCTTCTGCTCACAGTAATAGCCACAGCATTTATAGGATACGTCCTACCATGAGGACAAATATCATTCTGAGGAGCAACAGT
CATCACCAACCTCTTATCAGCAATCCCATACATCGGCAACAAATTTAGTCGAATGAATCTGAGGCGGATTCTCAGTAGACAAAGCAA
CCCTTACCCGATTCTTCGCTTTCCATTTTATCCTTCCATTTATCATCATAGCAATTGCCATAGTCCACCTACTATTCCTCCACGAAACA
GGCTCCAACAACCCAACAGGAATTTCCTCAGACGTAGACAAAATCCCATTCCACCCCTACTATACCATTAAGGACATCTTAGGGG
CCCTCTTACTAATTCTAGCTCTAATACTACTAGTACTATTCGCACCCGACCTCCTCGGAGACCCAGATAACTACACCCCAGCCAATC
CACTCAACACACCCCCTCACATCAAACCCCGAGTGATACTTCTTATTTGCATACGCAATCTTACGATCAATCCCCAACAAACTAGGA
GGAGTACTAGCCCTAGCCTTCTCTATCCTAATTCTTGCTCTAATCCCCTACTACACACCTCCAAACAACGAAGCATAATATTCCGA
CCACTCAGCCAATGCCTATTCTGAGCCCTAGTAGCAGACCTATTGACACTCACATGAATTGGAGGACAACCAGTCGAACACCCA
TATATCACCATCGGACAACTAGCATCTGTCCTATACTTTCTCCTCATCCTAGTGCTAATACCAACGGCCGGCACAATCGAAAACAAA
TTACTAAAATGAAGA···
>seq2
ATG···
ATCAACATCCGAAAAACCCACCCACTAATAAAAATTGTAAACAACGCATTCATTCATTCCCCAGCTCCATCAAATATTTCATCATGAT
GAAACTTTGGCTCTCTCCTAGGCATTTGCTTAATTTTACAGATTCTAACAGGCCTATTCCTAGCAATACACTATACACCTGACAA
CAACAGCATTCTCCTCTGTAACCCACATTTGCCGAGACGTAAACTATGGCTGAATTATCCGATATACACGCAAACGGGGCATCA
ATATTTTTTATCTGCCTATTTATGCATGTAGGACGAGGCCTATACTATGGATCATATACCTTCCTAGAAACATGAAACATCGGCATCA
```

<- Back

Introduction | The Java TreeBuilder application | The database behind the web site | The DNA Surveillance (Admin) web site
○○ ○○ ○○ ○○○

Appendix

DNA Surveillance Administration)

Species identification with DNA

Edit sequence details to alignment ' References '

Set source specimen and domain for every sequence in the alignment.

[Show current alignment]

Current reference alignments

sequence name	specimen	domain
seq1	seq1 ▾	Mammal ▾
seq2	seq2 ▾	Mammal ▾
seq3	seq3 ▾	Mammal ▾
seq4	seq4 ▾	Mammal ▾
seq5	seq5 ▾	Mammal ▾
seq6	seq1 ▾	Mammal ▾
seq7	seq1 ▾	Mammal ▾
seq8	seq1 ▾	Mammal ▾
seq9	seq1 ▾	Mammal ▾
seq10	seq1 ▾	Mammal ▾
seq11	seq1 ▾	Mammal ▾

[Submit] [Cancel]

▸ back