Introduction to **wEMBOSS** (EMBOSS)

Shahid Manzoor
Adnan Niazi

SLU Global Bioinformatics Centre, Uppsala, Sweden

What is  ?

- A free Open Source software analysis package developed for molecular biology
- Programs share a common look and feel
- Incorporates many small and large programs
- Easy to run from the command line
- Retrieval of sequence data from the web
- Easy to call from other programs(e.g. perl)
- Easy to set up behind GUIs and Web interfaces

Scope of applications

- There are many EMBOSS programs (200+)
- Many sequence analysis & display programs
- Protein 3D structure prediction being developed
- Other programs e.g enzyme kinetics

What is wEMBOSS?

- A web interface to the EMBOSS package for sequence analysis

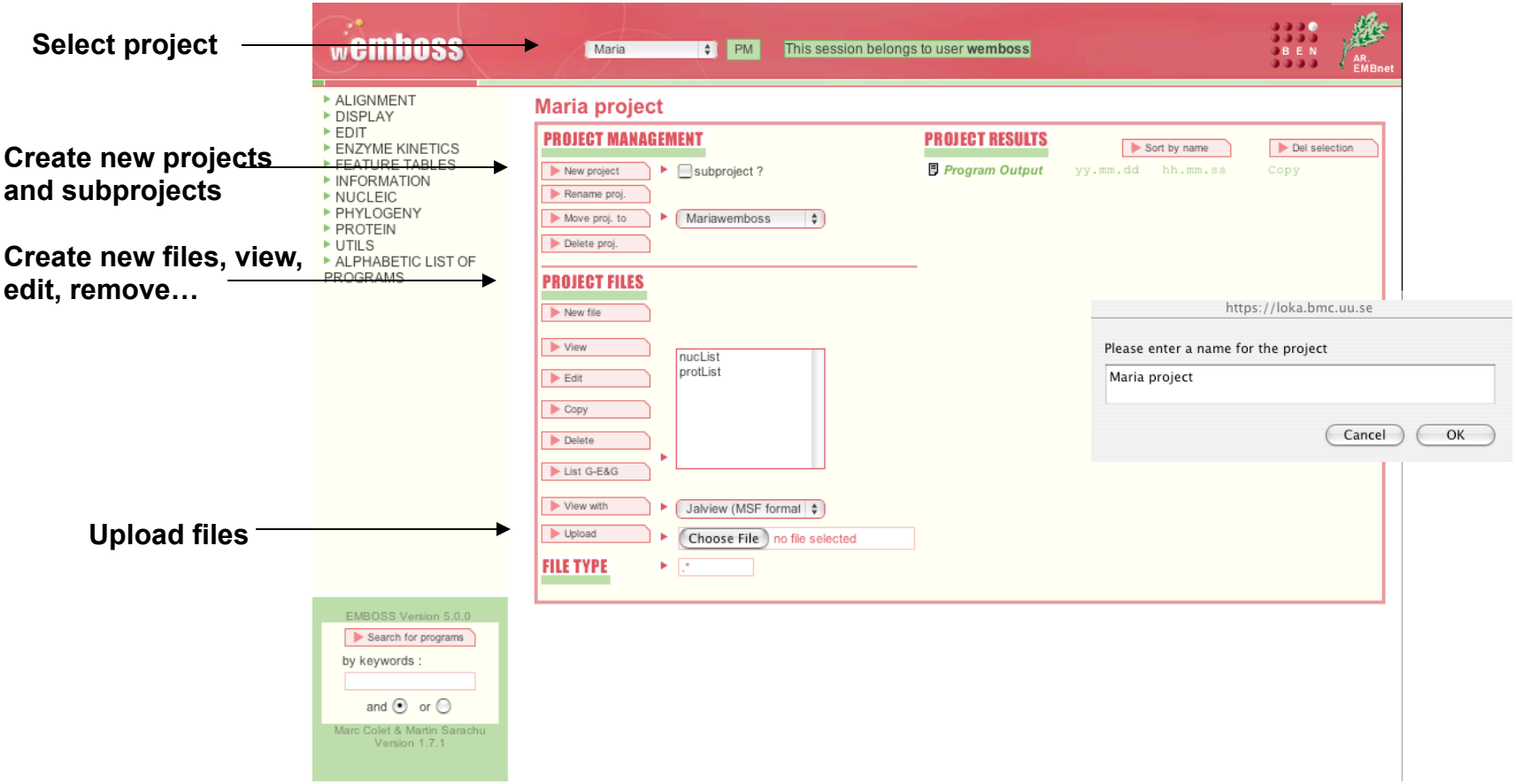
- It's developed by Martin Sarachu (Argentina) and Marc Colet (Belgium)

Feautures of wEMBOSS:

- Each user has a separate and private workspace.
- Organize your work by creating projects and subprojects
- Results saved for easy recover & review
- Inline help
- Keyword search for programs and documentation.

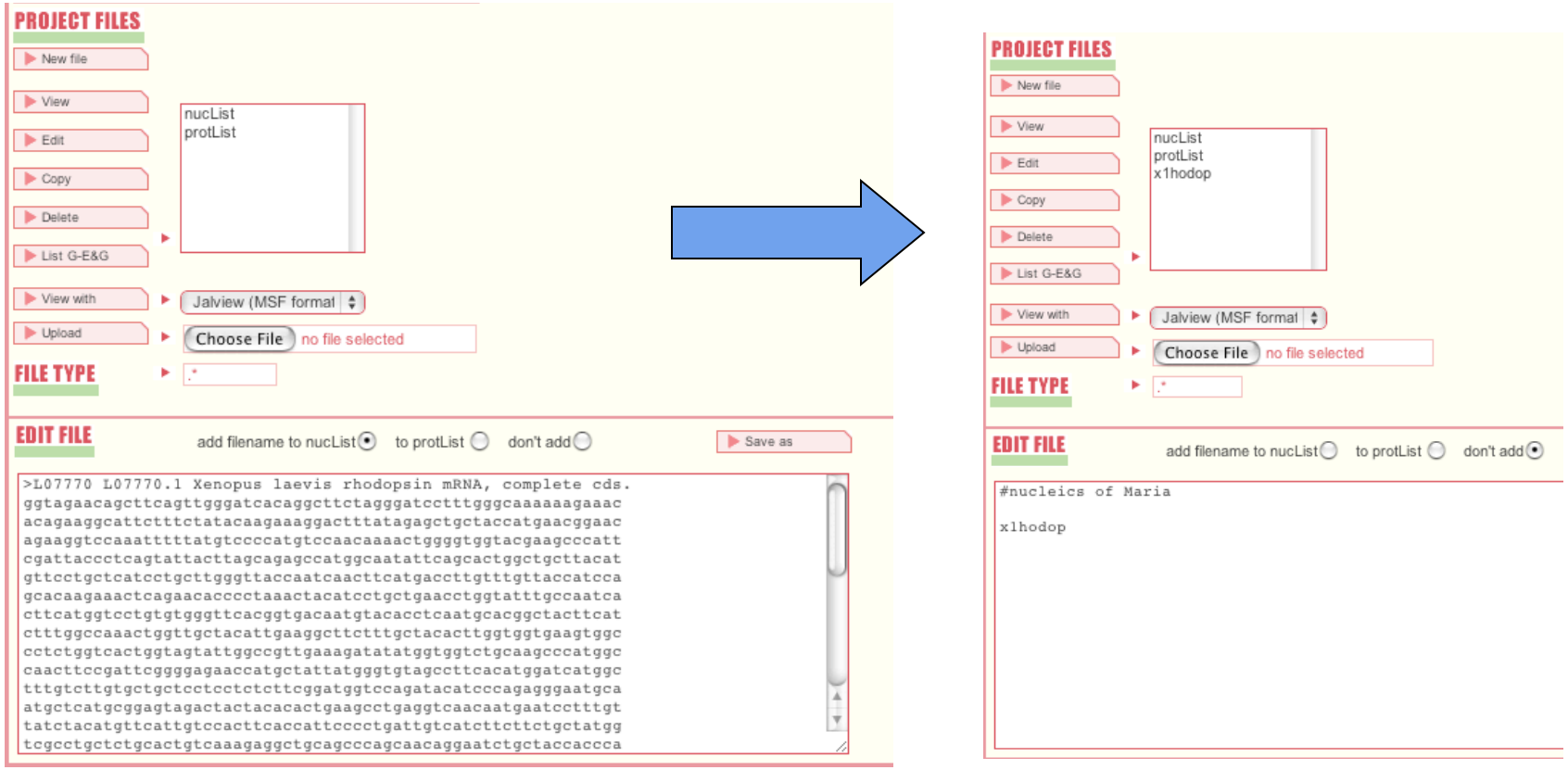
Create a project

- **wEMBOSS** uses the concept of projects
- nucList & protList are automatically created (add names of the sequences you wish to access)



Project files

- Create a new file and paste the sequence or upload sequence
- Add sequence name to nucList or protList



Selection of programs/files

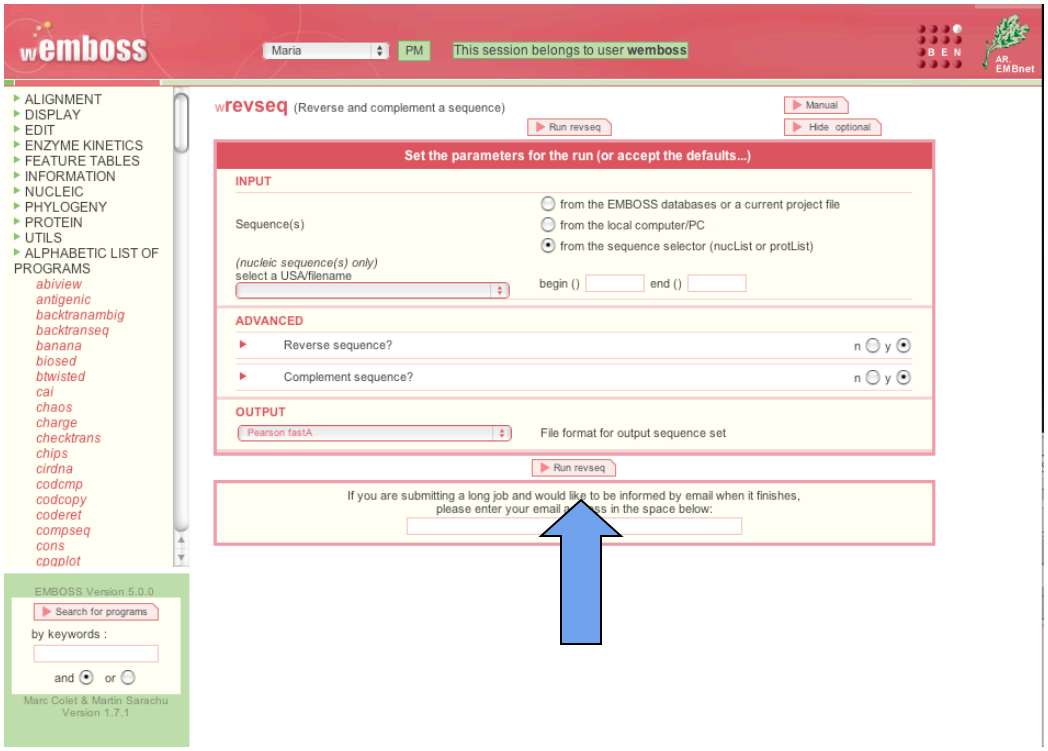
- Drop-down menu with all available programs
- Select a program by clicking on its name.

- Choose a sequence to work with from:

  - list selector: to select a sequence from nucList or protList

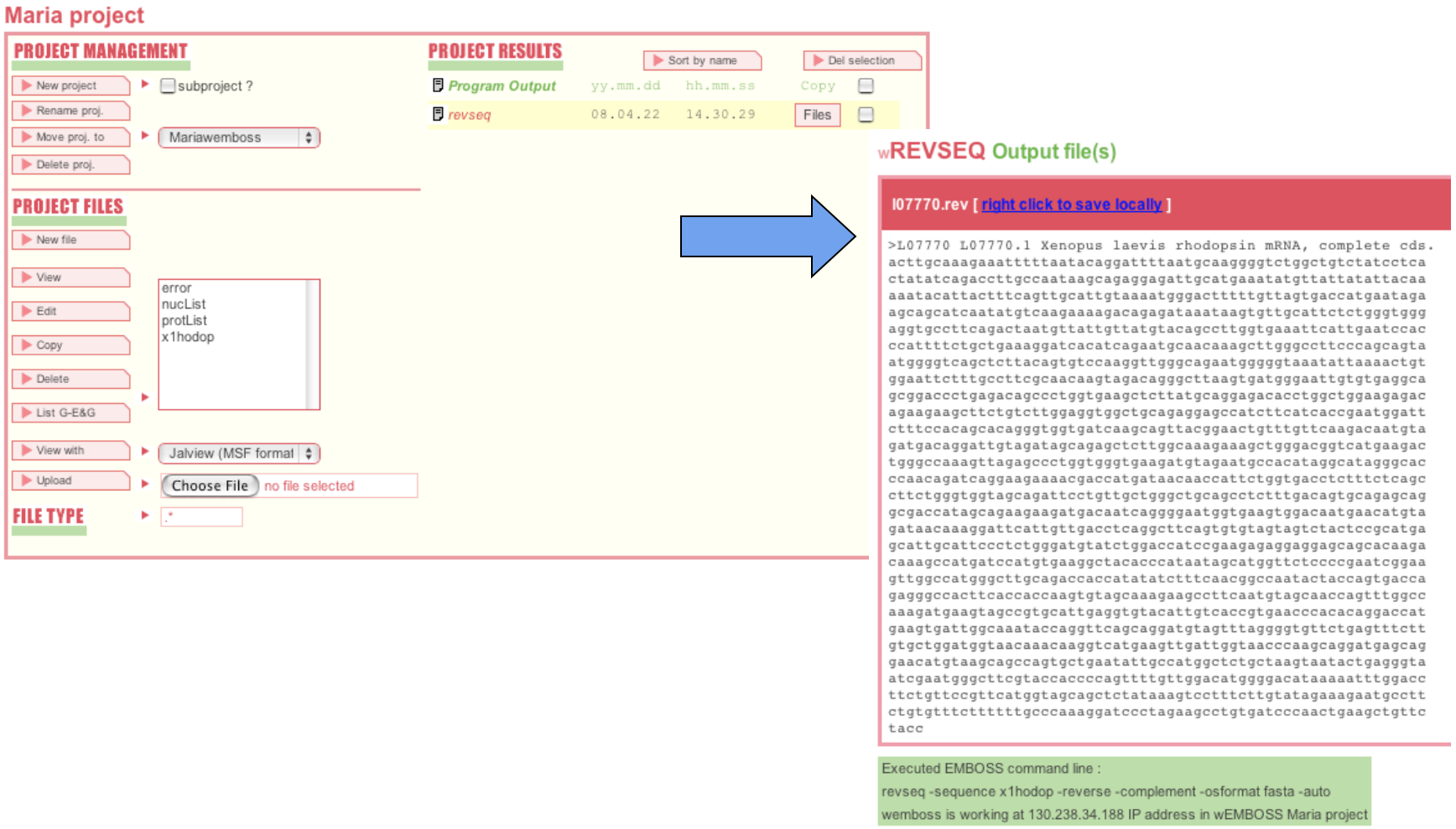
  - local computer file: to upload a file from your computer

  - EMBOSS database: to access a sequence from a server

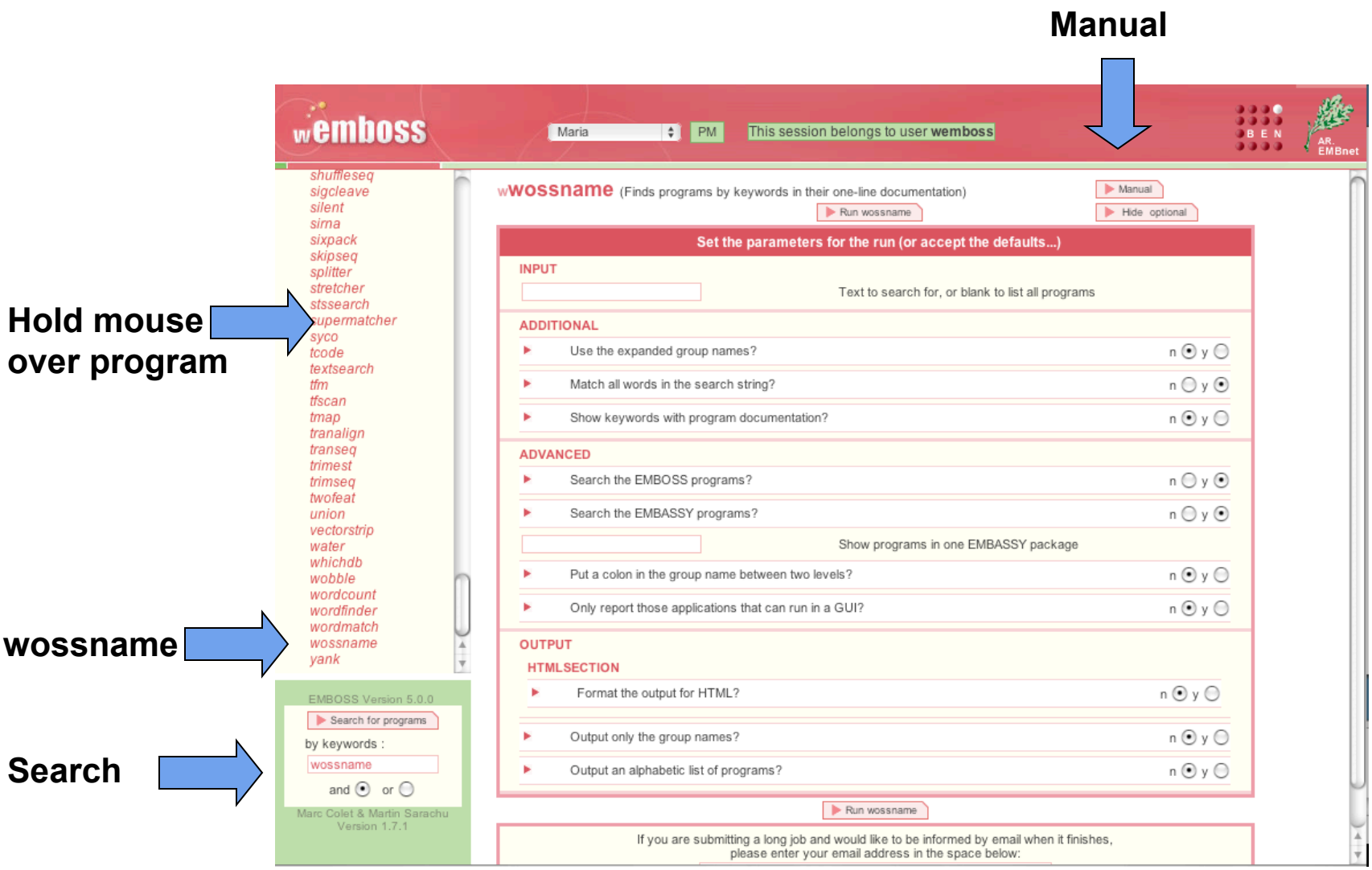


Project results

- Results automatically viewed and saved

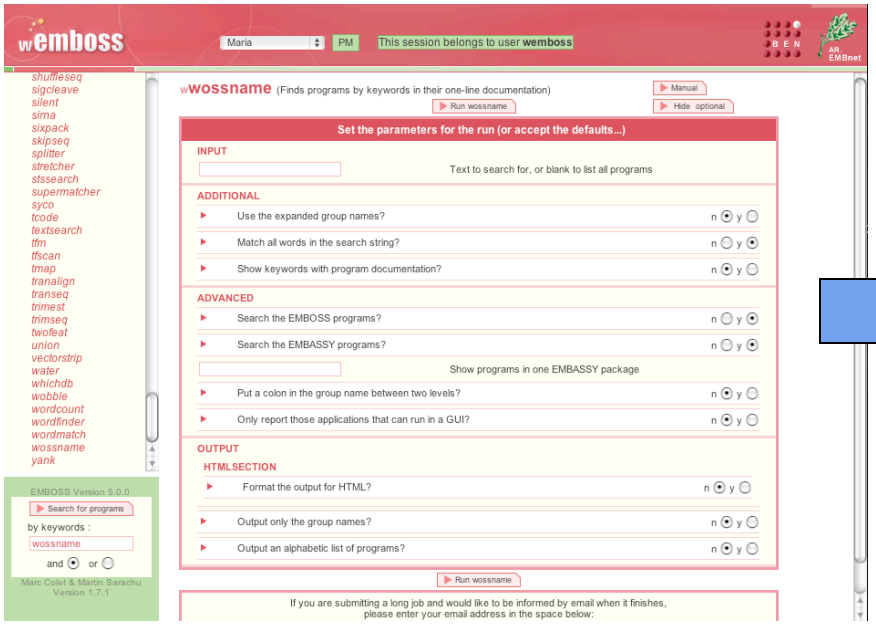


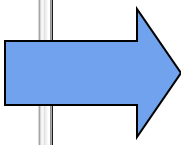
Help within wEMBOSS:



wEMBOSS program: **wossname**

- Easy to forget a program name
- To find programs, use **wossname**
- **wossname** finds programs by looking for keywords in the description or the name of the program





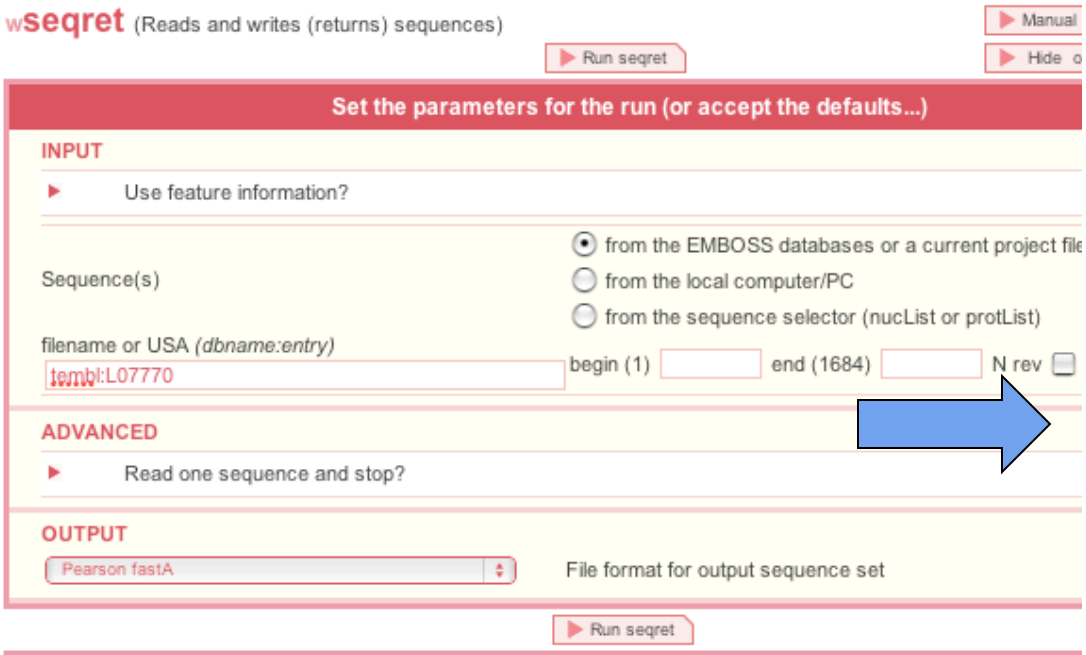
wWOSSNAME Output file(s)

Warning

```
SEARCH FOR 'PROTEIN'
antigenic       Finds antigenic sites in proteins
backtranambig   Back translate a protein sequence to ambiguous codons
backtranseq     Back translate a protein sequence
charge          Protein charge plot
checktrans      Reports STOP codons and ORF statistics of a protein
compseq         Count composition of dimer/trimer/etc words in a sequence
digest          Protein proteolytic enzyme or reagent cleavage digest
emowse          Protein identification by mass spectrometry
epestfind       Finds PEST motifs as potential proteolytic cleavage sites
fproml          Protein phylogeny by maximum likelihood
fpromlk         Protein phylogeny by maximum likelihood
fprotdist       Protein distance algorithm
fprotpars       Protein parsimony algorithm
freak           Residue/base frequency table or plot
fuzzpro         Protein pattern search
fuzztran        Protein pattern search after translation
garnier         Predicts protein secondary structure
helixturnhelix  Report nucleic acid binding motifs
hmoment         Hydrophobic moment calculation
iep             Calculates the isoelectric point of a protein
makeprotseq     Creates random protein sequences
msbar           Mutate sequence beyond all recognition
mwcontam        Shows molwts that match across a set of files
mwfilter        Filter noisy molwts from mass spec output
octanol         Displays protein hydropathy
oddcomp         Find protein sequence regions with a biased composition
patmatdb        Search a protein sequence with a motif
patmatmotifs    Search a PROSITE motif database with a protein sequence
pepcoil         Predicts coiled coil regions
pepinfo         Plots simple amino acid properties in parallel
pepnet          Displays proteins as a helical net
pepstats        Protein statistics
pepwheel        Shows protein sequences as helices
pepwindow       Displays protein hydropathy
pepwindowall    Displays protein hydropathy of a set of sequences
preg            Regular expression search of a protein sequence
profit          Scan a sequence or database with a matrix or profile
prophecy        Creates matrices/profiles from multiple alignments
prophet         Gapped alignment for profiles
pscan           Scans proteins using PRINTS
psiphi          Phi and psi torsion angles from protein coordinates
shuffleseq      Shuffles a set of sequences maintaining composition
```

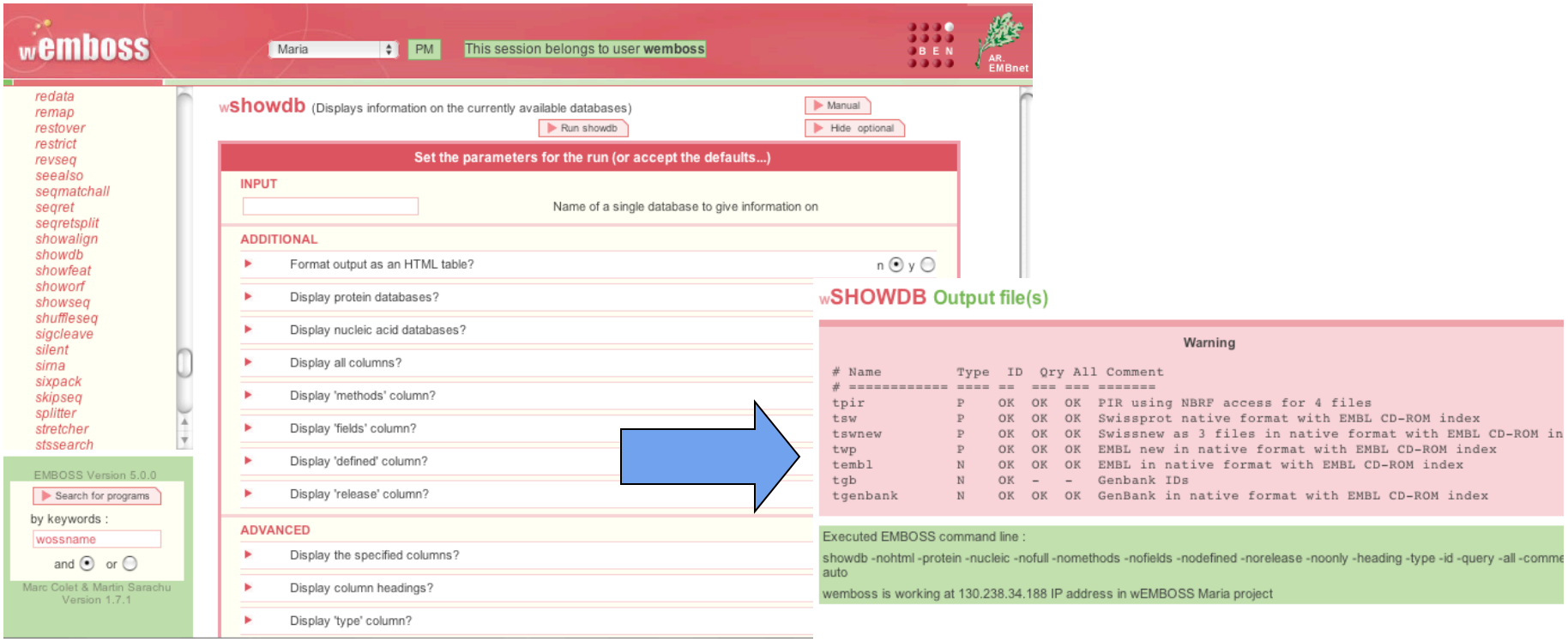
wEMBOSS program: **seqret**

- Reads in a sequence and writes it out
- Reformat sequences
- Get sequences from databases



wEMBOSS program: **showdb**

- Displays information of the currently available databases

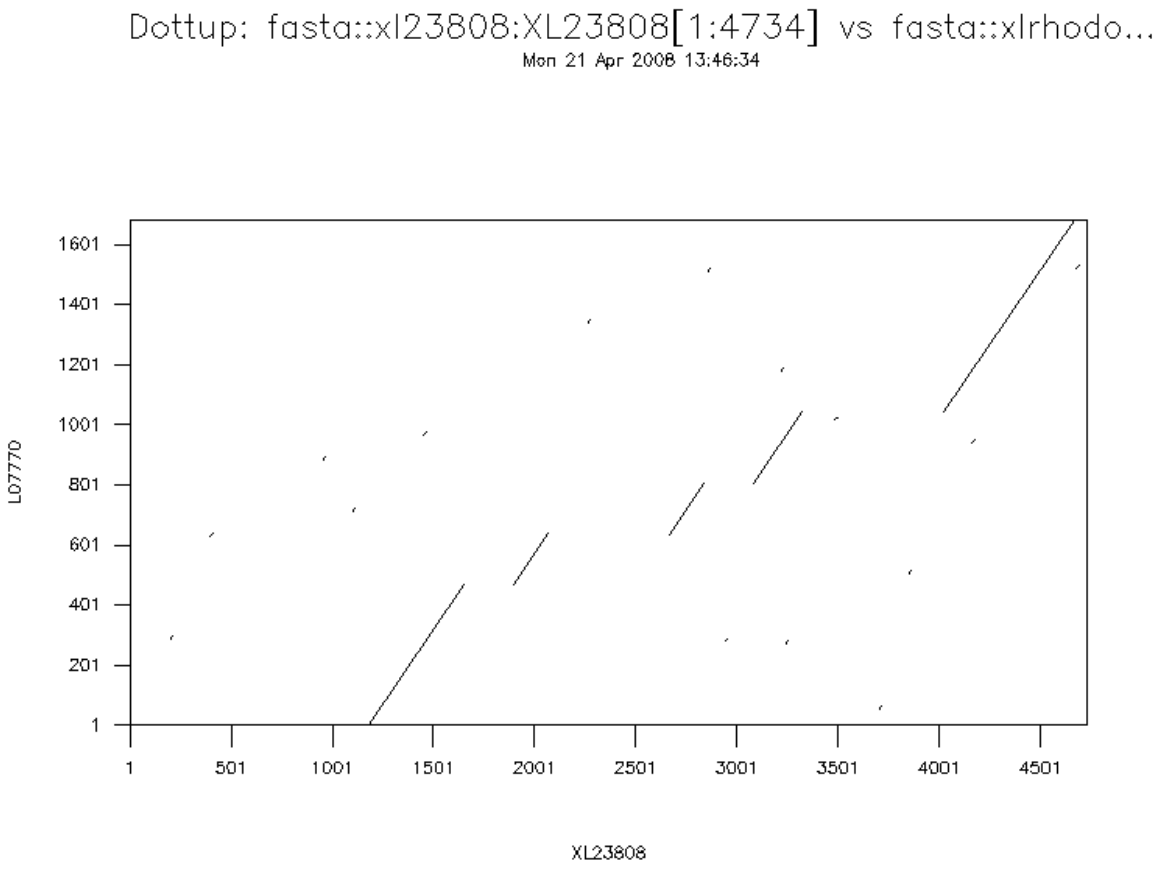


Examples of other wEMBOSS programs

- Pairwise alignment -Dotup
- Global and local alignment – needle, water
- Protein translation – plotorf, getorf, transeq
- Protein fingerprint - pscan
- Primers - eprimer3
- Microsatellites -equicktandem

Pairvise alignment - Dotup

- One sequence is represented on each axis and significant matching regions are distributed along diagonals in the matrix.



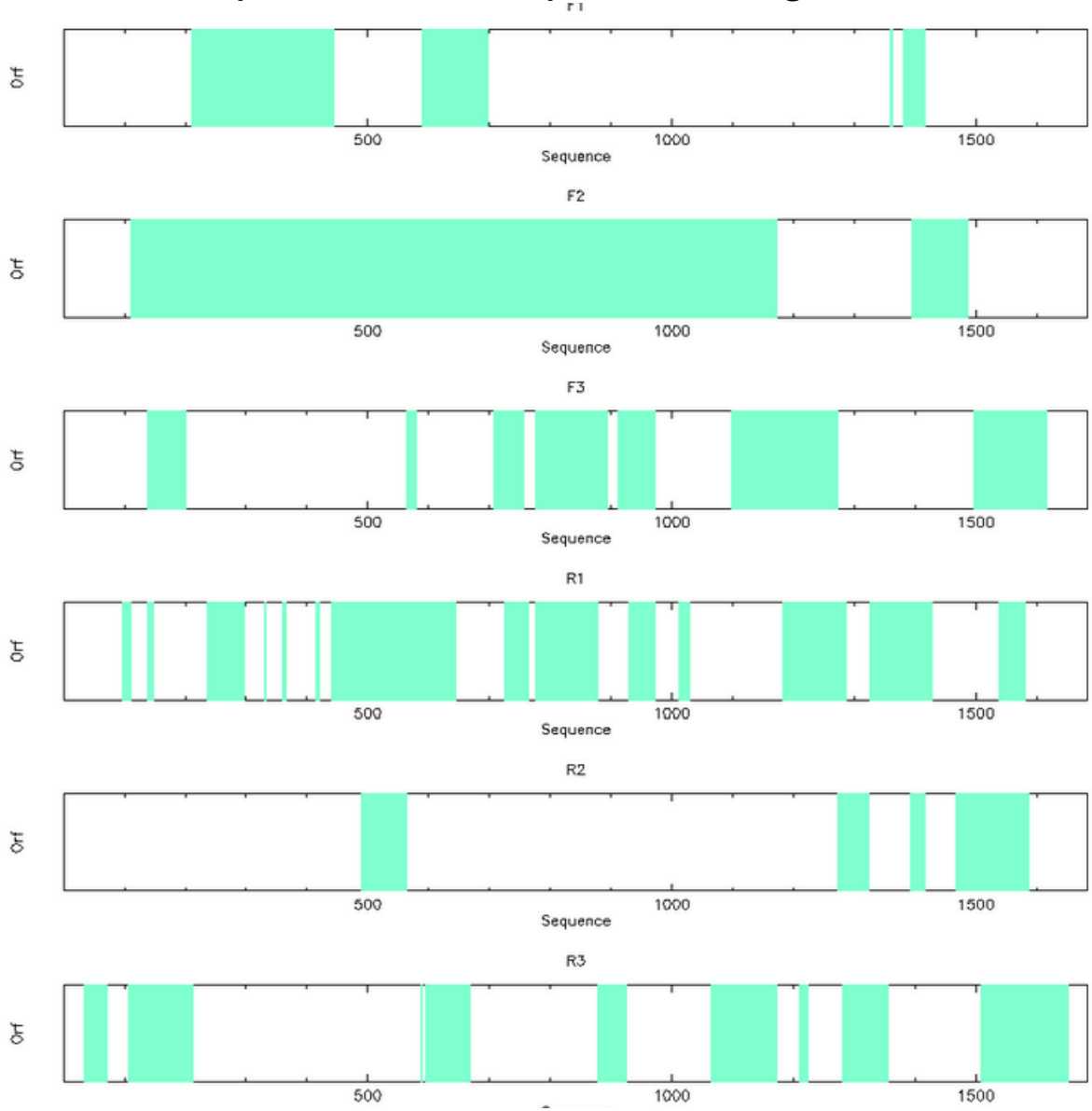
wEMBOSS program: **water**

- Does a local alignment of two sequences

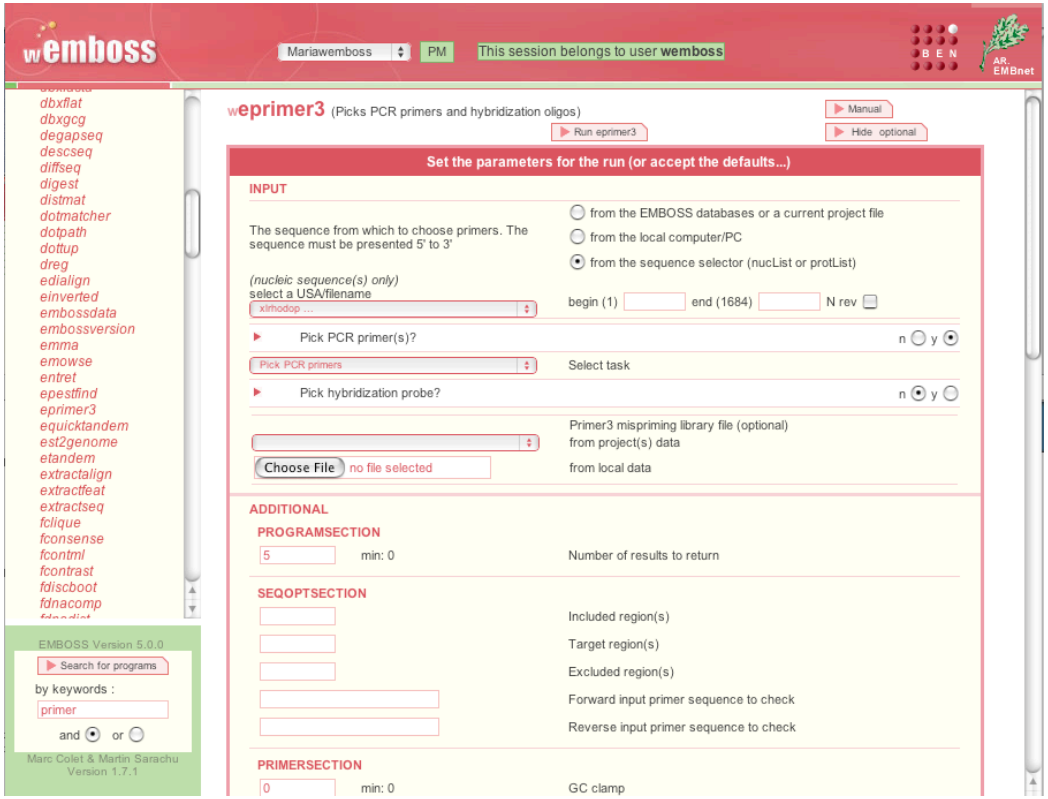
```
L07770          88 aggactttatagagctgctaccatgaacggaacagaaggtccaaattttt  137
                   |||||||||.|||||.||.|..||||||||||||||||.|||||.||||
BMU59922        10 AGGACTTTAAAGAGCCGCCAATATGAACGGAACAGAAGGCCCAAACTTTT   59

L07770         138 atgtccccatgtccaacaaaactggggtggtacgaagcccattcgattac  187
                   |..|.|||||||||||||.|||||||||||||||||||||||||.|||
BMU59922        60 ACATACCCATGTCCAACAAGACTGGGGTGGTACGAAGCCCATTCGAATAC  109

L07770         188 cctcagtattacttagcagagccatggcaatattcagcactg-gctgctt  236
                   ||||||||||||.|.||||||||||||||||.||.|...|| || ||.|
BMU59922       110 CCTCAGTATTACCTGGCAGAGCCATGGCAATACTCCGTTTTGTGC-GCAT  158

L07770         237 acatgttcctgctcatcctgcttgggttaccaatcaacttcatgaccttg  286
                   ||||||||||||||.||.||||||||.||||||||||||||||||||||
BMU59922       159 ACATGTTCCTGCTCATTCTTCTTGGGTTCCCAATCAACTTCATGACCTTG  208

L07770         287 tttgttaccatccagcacaagaaactcagaacacccctaaactacatcct  336
                   |..||.||||||||||||||||.|||.|.||||||.|||||||.|||||
BMU59922       209 TACGTCACCATCCAGCACAAGAAGCTCCGGACACCCTTAAACTATATCCT  258

L07770         337 gctgaacctggtatttgccaatcacttcatggtcctgtgtgggttcacgg  386
                   ||||||..|||..||||||||.||||||||||||||||||.|||||.|
BMU59922       259 GCTGAATTTGGCCTTTGCCAACCACTTCATGGTCCTGTGTGGATTCACCG  308

L07770         387 tgacaatgtacacctcaatgcacggctacttcatctttggccaaactggt  436
                   |.||||||||.|||||||.||||.||||||.||||.||.|||||.|||
BMU59922       309 TCACAATGTACTCCTCAATGAACGGATACTTCGTCTTCGGACAAACCGGT  358

L07770         437 tgctacattgaaggcttctttgctacacttggtggtgaagtggccctctg  486
                   |||||..||||||||||||.|||||.||||||||||||.|.|||||.||
BMU59922       359 TGCTATGTTGAAGGCTTCTTCGCTACCCTTGGTGGTGAAATCGCCCTTTG  408
```

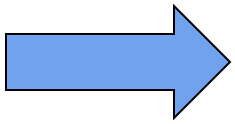
wEMBOSS program: **plotorf**

- Translates sequences to 6 open reading frames.



Pick primers -eprimer3





wEPRIMER3 Output file(s)

l07770.eprimer3 [right click to save locally]

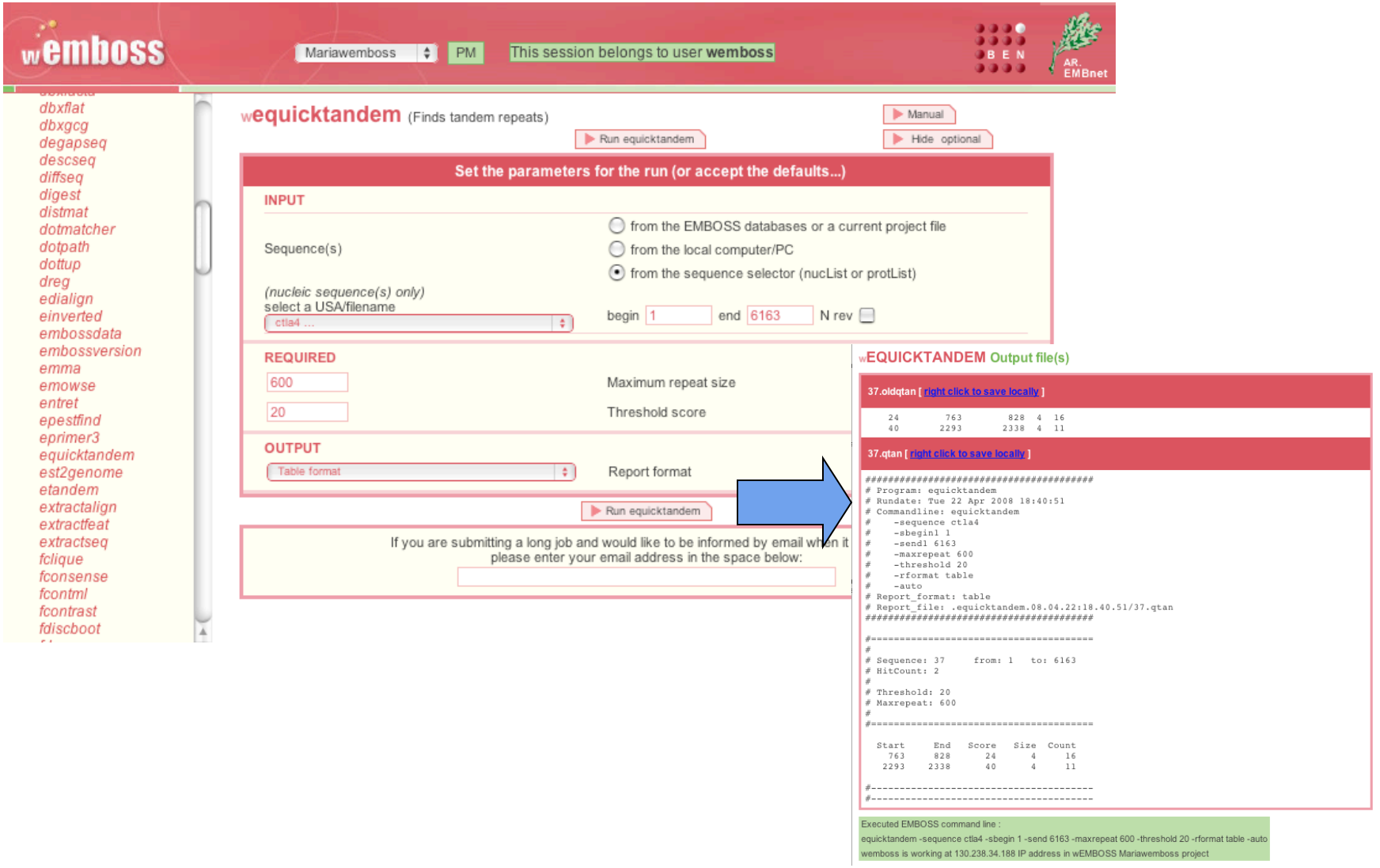
```
# EPRIMER3 RESULTS FOR L07770

#                      Start  Len    Tm     GC%   Sequence

   1 PRODUCT SIZE: 199
     FORWARD PRIMER      425   20  60.00  50.00  GGCCAAACTGGTTGCTACAT

     REVERSE PRIMER      604   20  60.14  60.00  GAGGAGGAGCAGCACAAGAC


   2 PRODUCT SIZE: 200
     FORWARD PRIMER      425   20  60.00  50.00  GGCCAAACTGGTTGCTACAT

     REVERSE PRIMER      605   20  60.28  55.00  AGAGGAGGAGCAGCACAAGA


   3 PRODUCT SIZE: 201
     FORWARD PRIMER      425   20  60.00  50.00  GGCCAAACTGGTTGCTACAT

     REVERSE PRIMER      606   20  60.28  60.00  GAGAGGAGGAGCAGCACAAG


   4 PRODUCT SIZE: 202
     FORWARD PRIMER      425   20  60.00  50.00  GGCCAAACTGGTTGCTACAT

     REVERSE PRIMER      607   20  60.28  55.00  AGAGAGGAGGAGCAGCACAA


   5 PRODUCT SIZE: 203
     FORWARD PRIMER      425   20  60.00  50.00  GGCCAAACTGGTTGCTACAT

     REVERSE PRIMER      608   20  60.28  55.00  AAGAGAGGAGGAGCAGCACA
```

Executed EMBOSS command line :

eprimer3 -sequence xlrhodop -primer -task 1 -nohybridprobe -numreturn 5 -gcclamp 0 -osize 20 -minsize 18 -maxsize 27 -otm 60.0 -mintm 57.0 -maxtm 63.0 -maxdifftm 100.0 -ogcpercent 50.0 -mingc 20.0 -maxgc 80.0 -saltconc 50.0 -dnaconc 50.0 -maxpolyx 5 -productosize 200 -productsizerange 100-300 -productotm 0.0 -productmintm -1000000.0 -productmaxtm 1000000.0 -noexplainflag -nofileflag -firstbaseindex 1 -nopickanyway -maxmispriming 12.00 -pairmaxmispriming 24.00 -numnsaccepted 0 -selfany 8.00 -selfend 3.00 -maxendstability 9.0 -auto

wemboss is working at 130.238.34.188 IP address in wEMBOSS Mariawemboss project

Find microsatellites - equicktandem



Start working on the tutorial