
A data management framework for the AFTOL project

Frank Kauff, Cymon J. Cox, Francois Lutzoni

Duke University, Durham, NC, USA

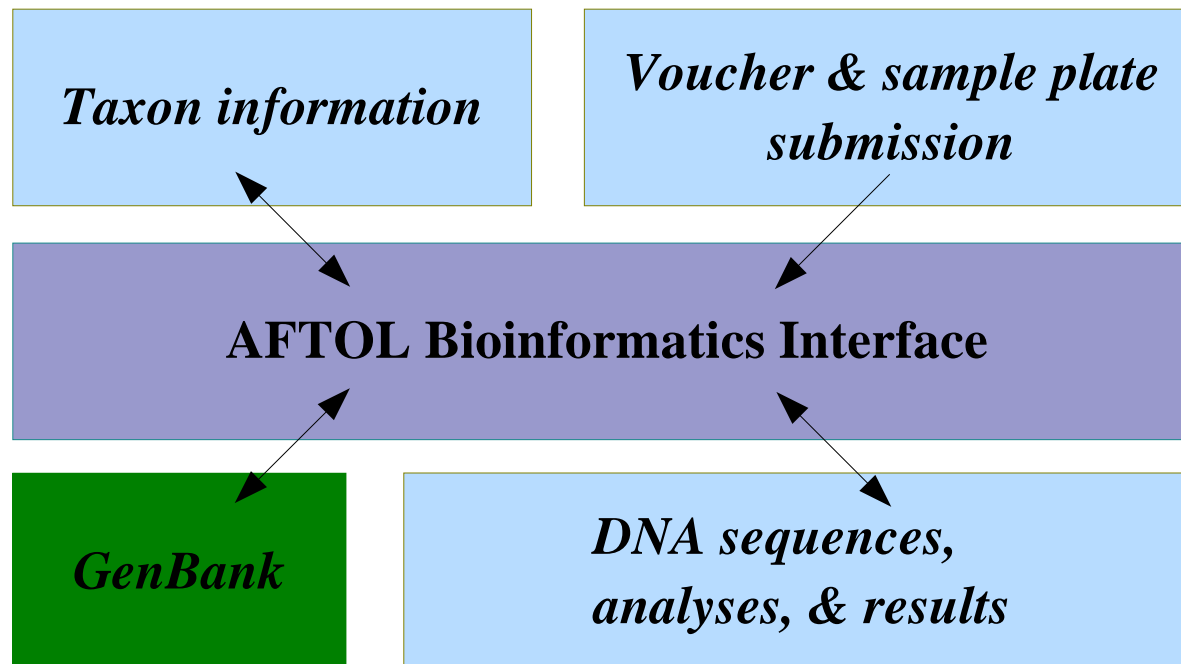
AFTOL – what, who and why is it?

- Assembling the Fungal Tree of Life, NSF funded – part of ATOL initiative
- Co-ordinating institutions: Clark University, Duke University, Oregon State University, University of Minnesota
- Aims to collect and analyze data from 1,500+ species of fungi for 8 genetic loci, plus a subset of taxa for morphological and ultrastructural characters

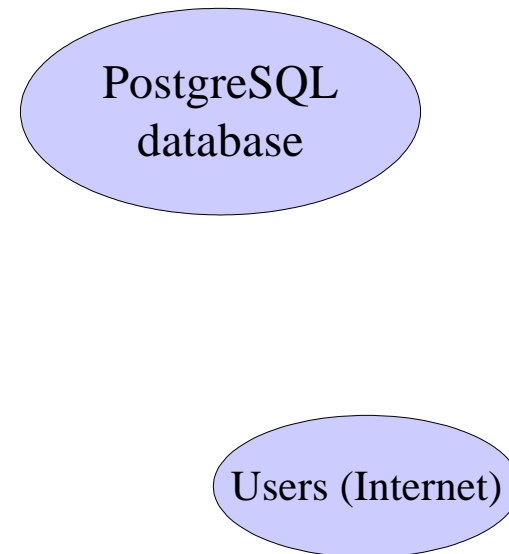


Provides:

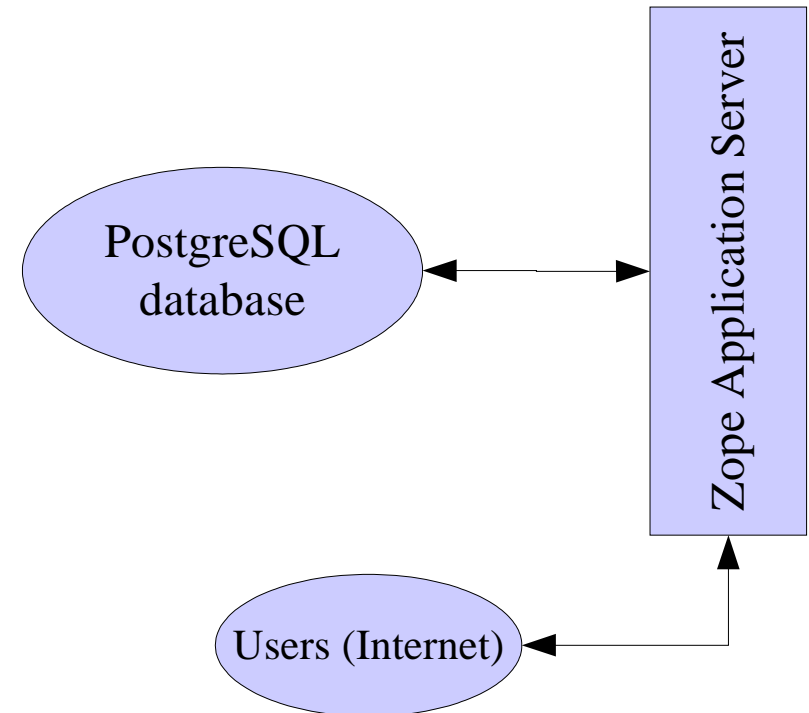
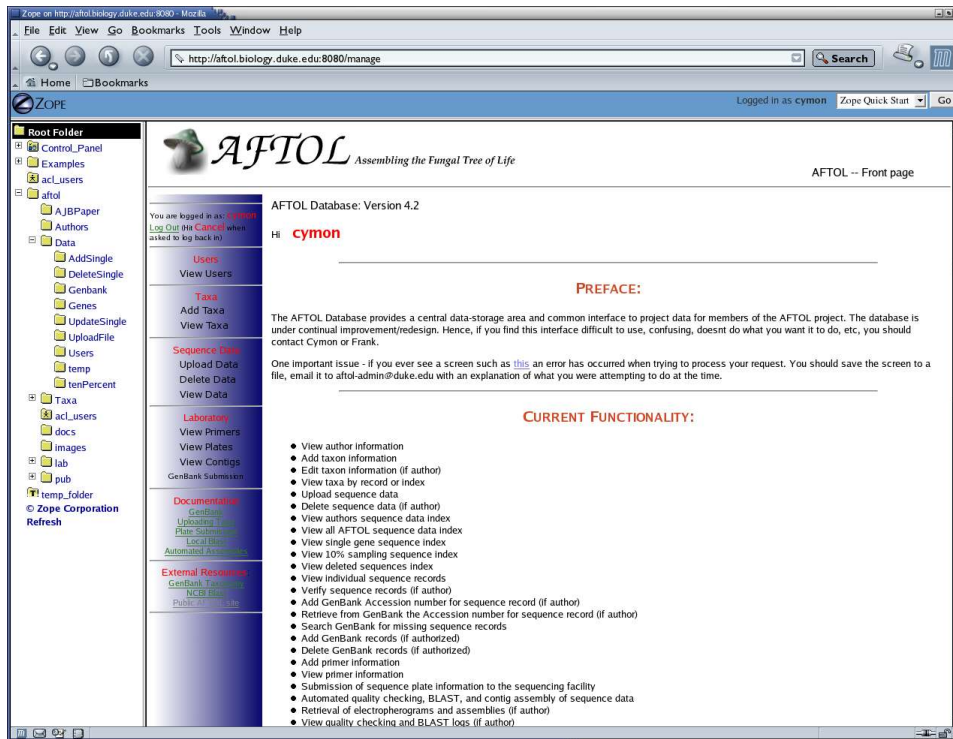
- Central storage for all project data
- Participant and public interface to the project data
- Automated analyses of raw sequence data: Phred, Phrap, local BLAST, ...
- Automated analyses of gene sequence data: alignment, congruence testing, and phylogenetic analyses



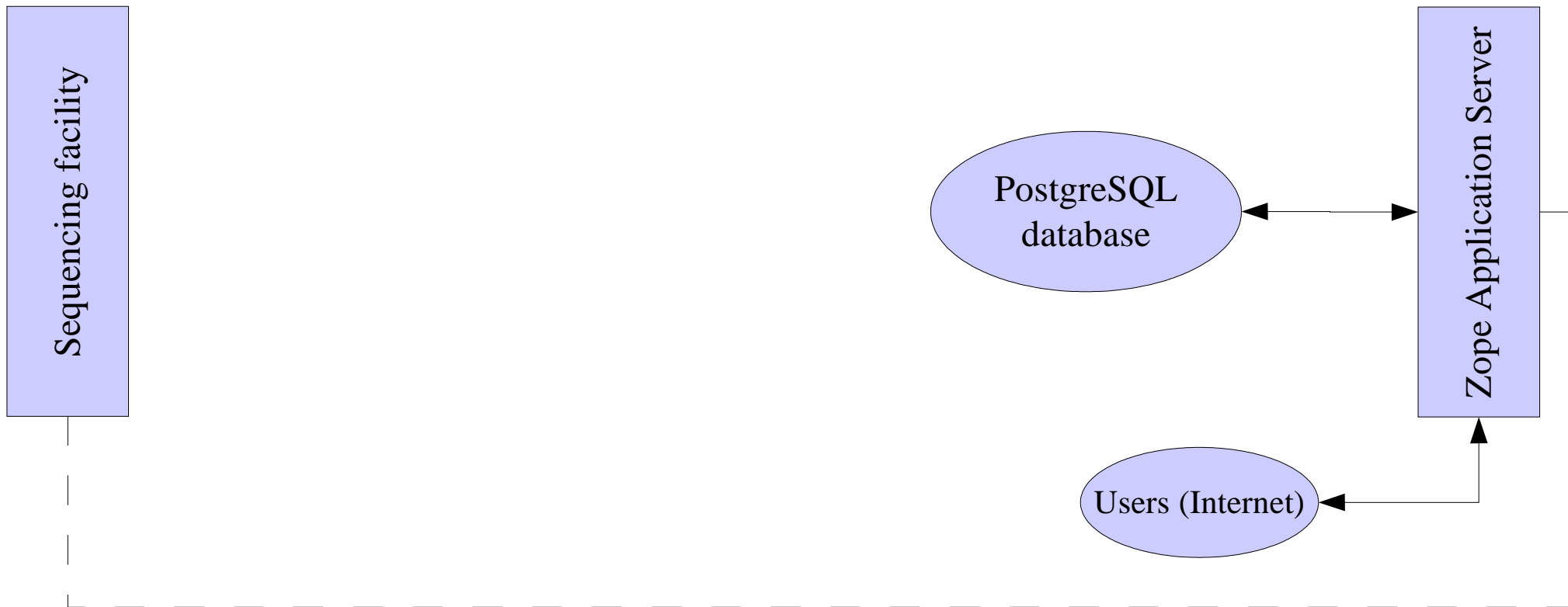
Overview of bioinformatic components



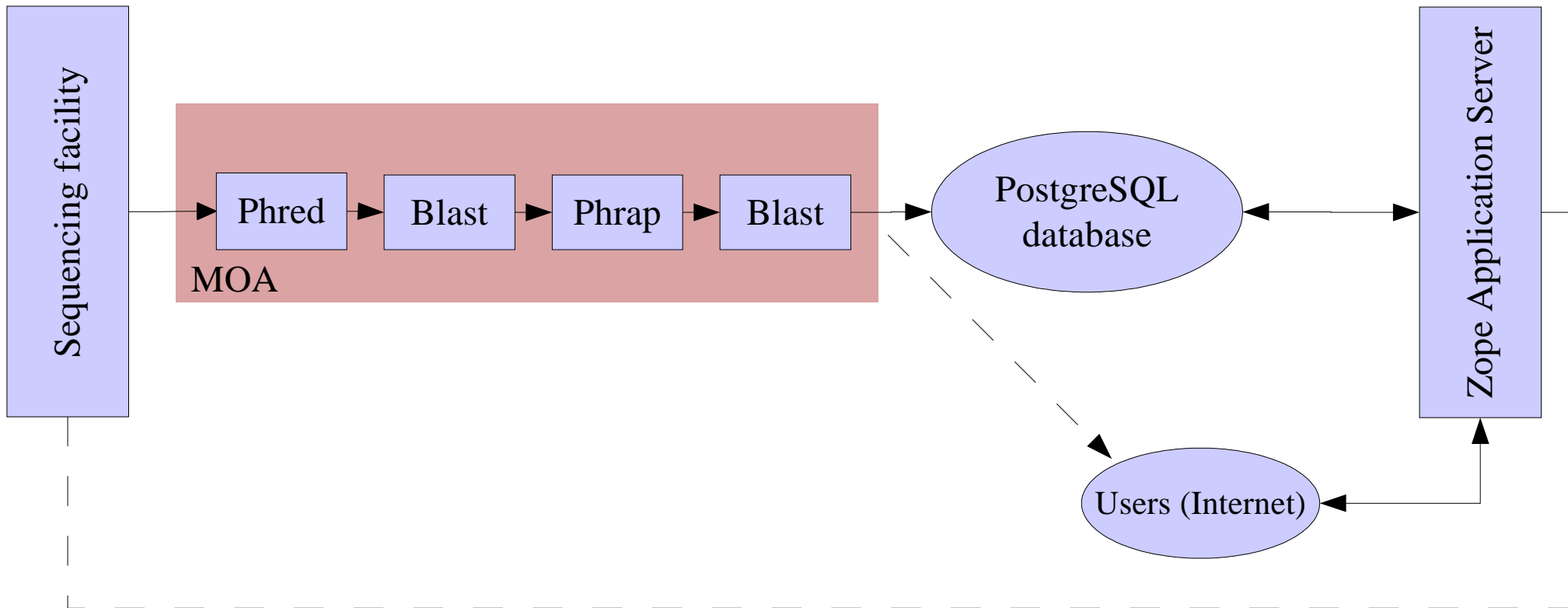
Overview of bioinformatic components



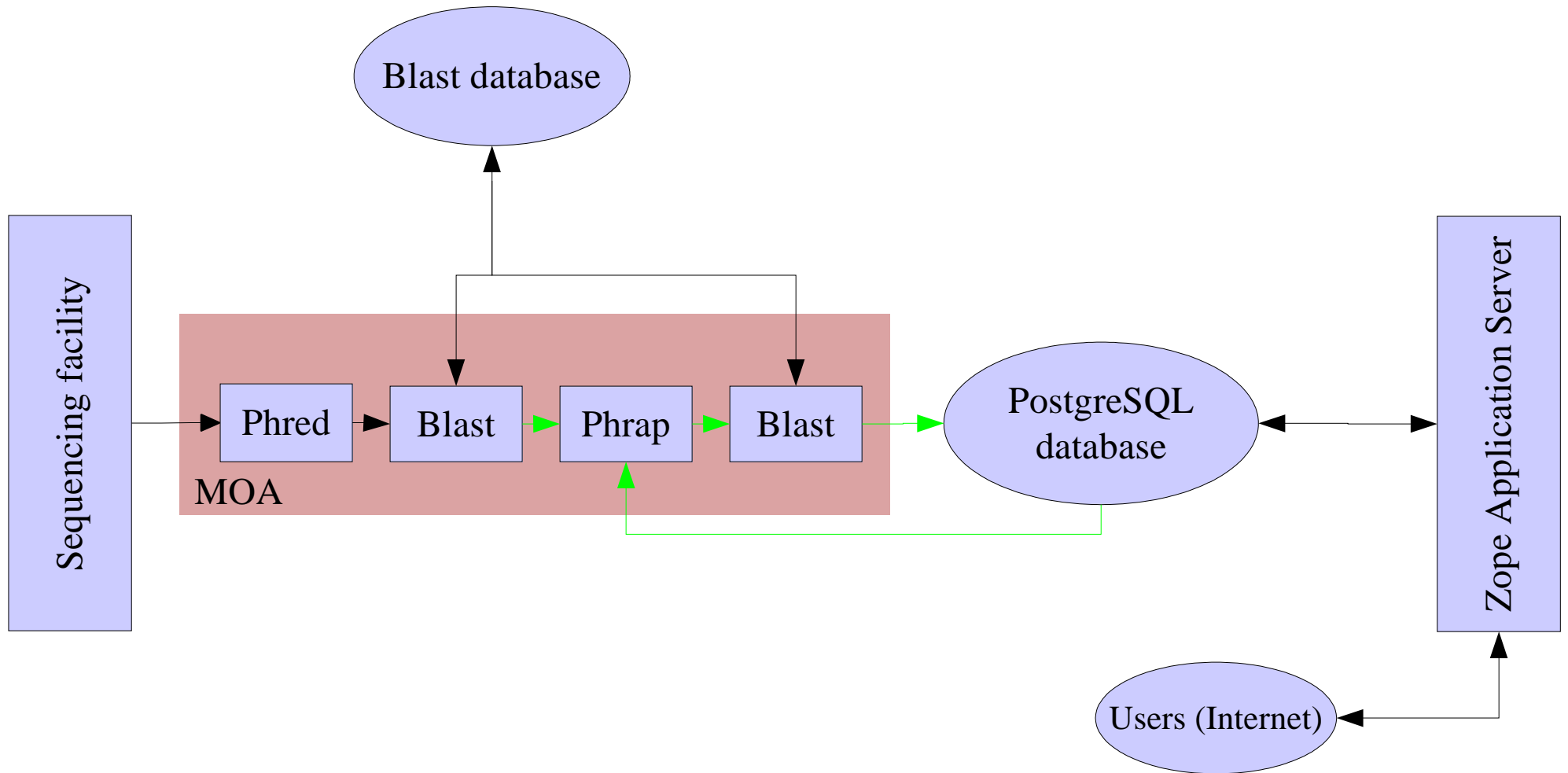
Overview of bioinformatic components



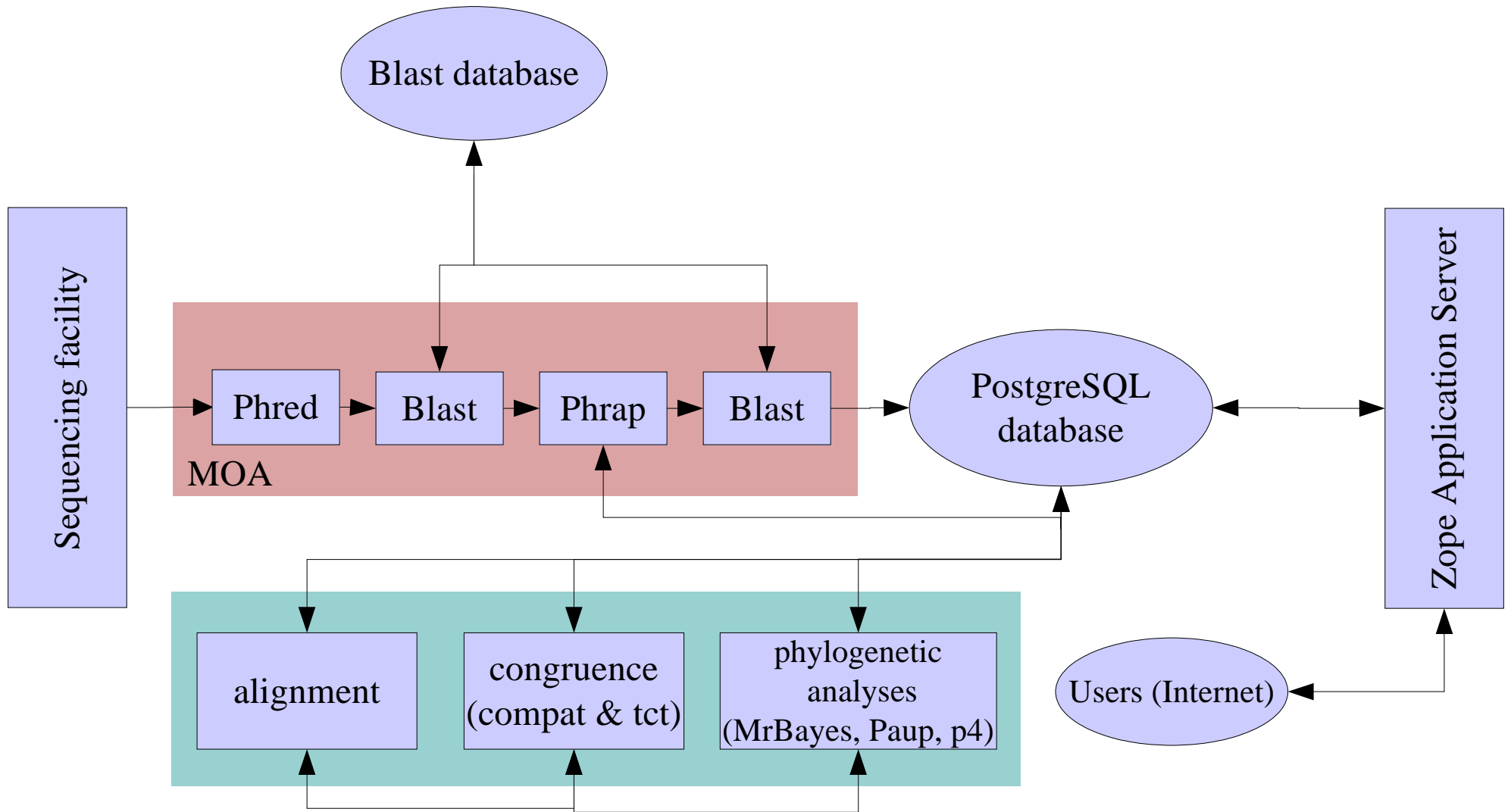
Overview of bioinformatic components



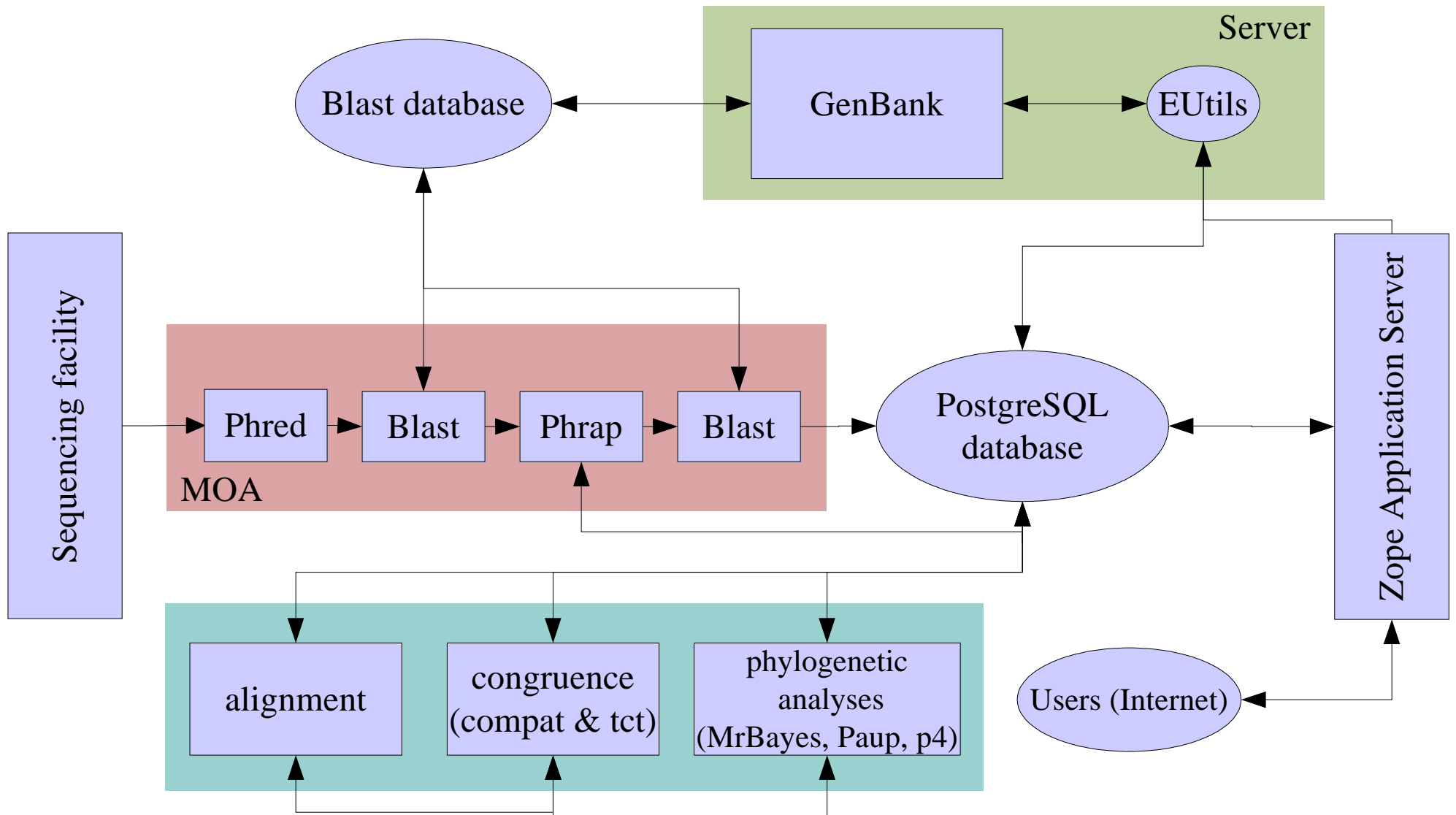
Overview of bioinformatic components



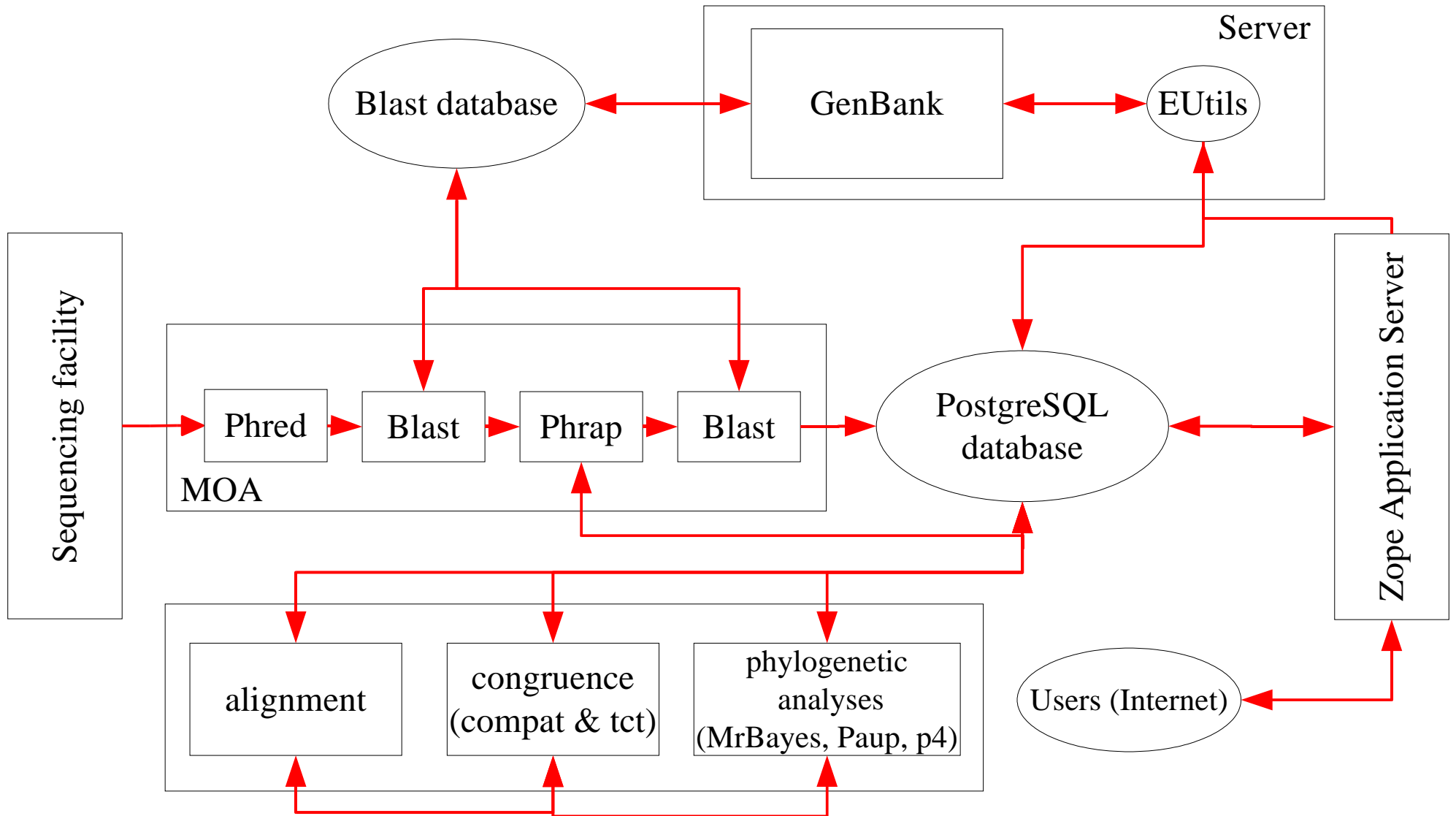
Overview of bioinformatic components



Overview of bioinformatic components

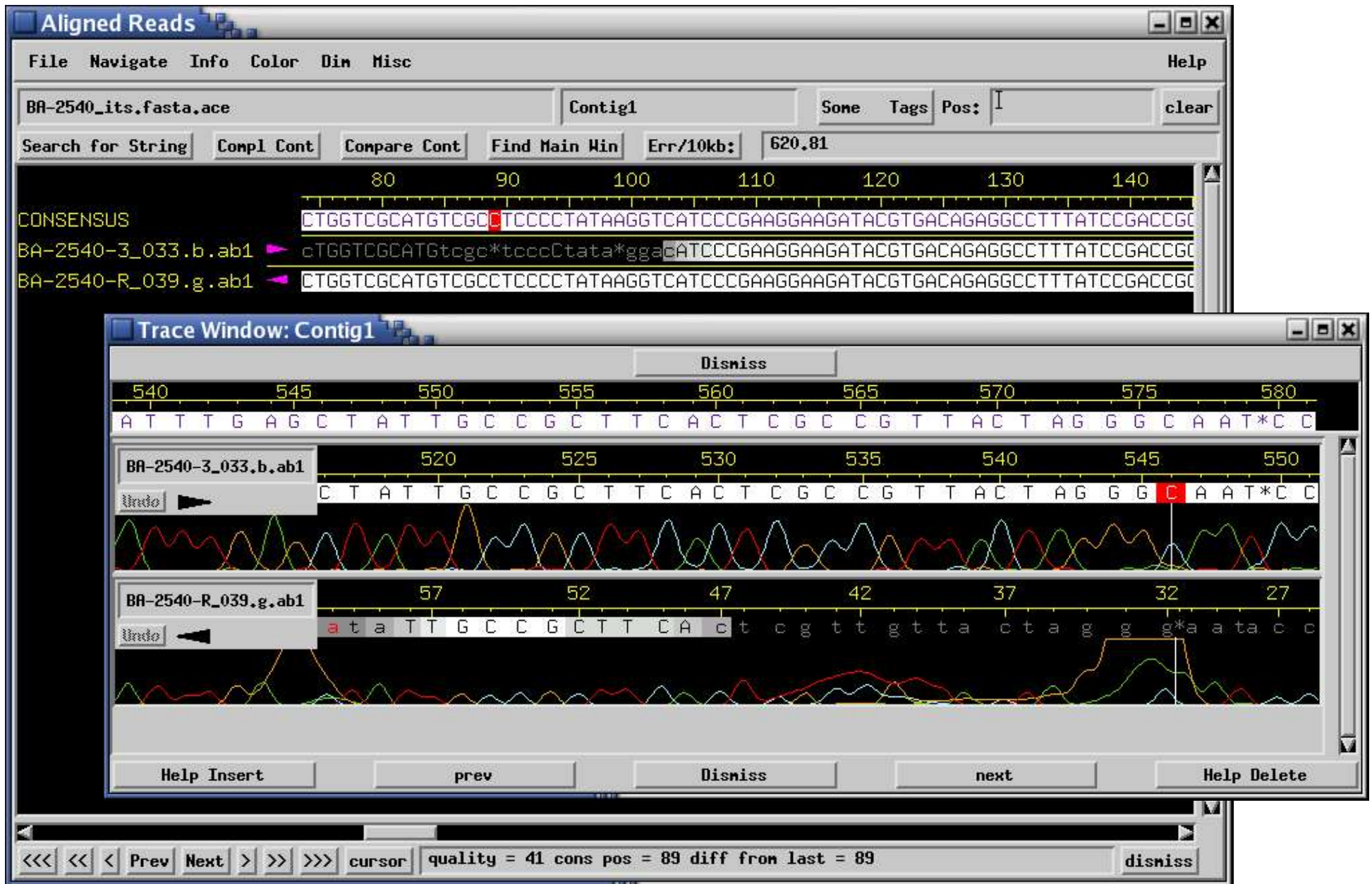


Overview of bioinformatic components

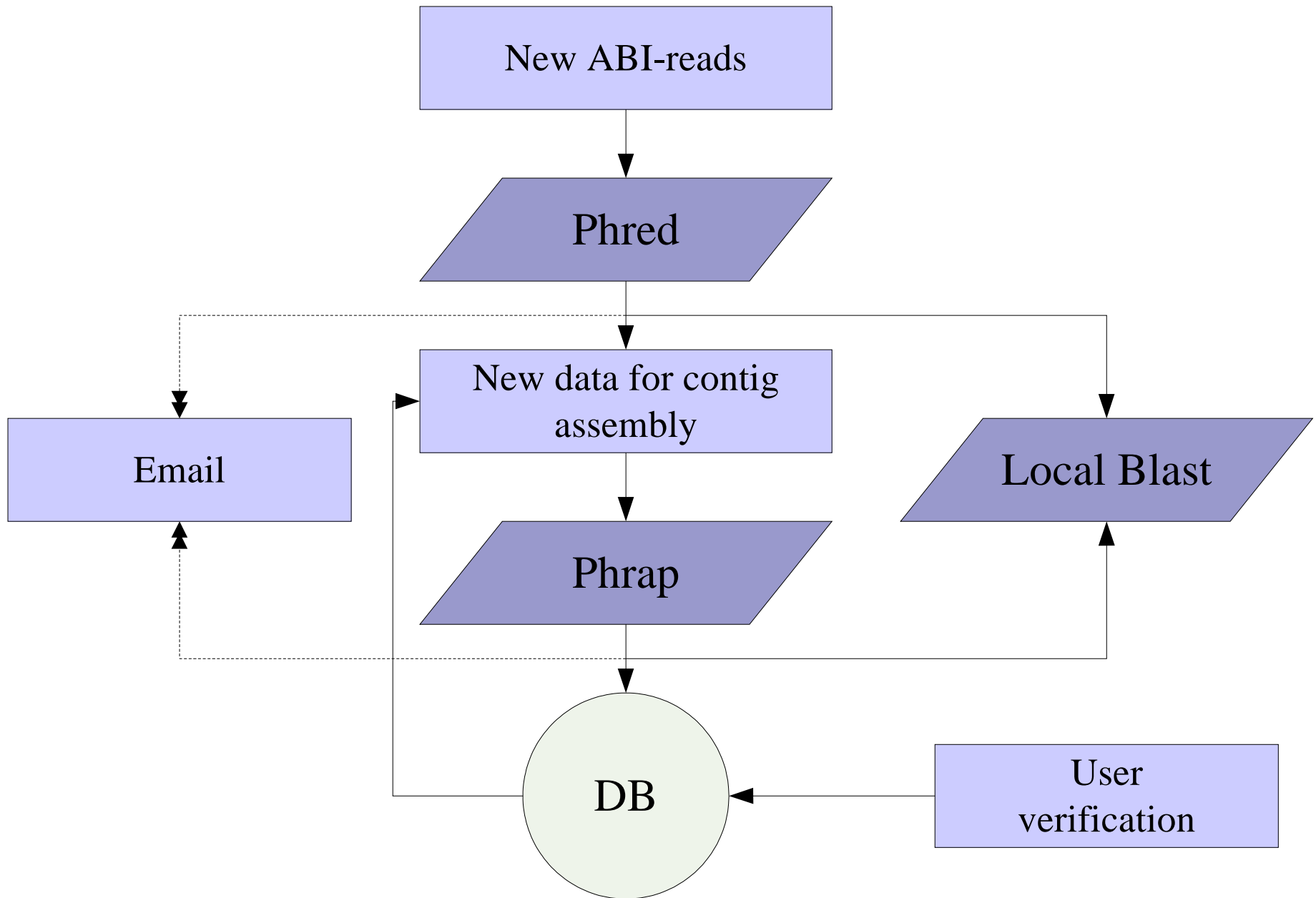


Python

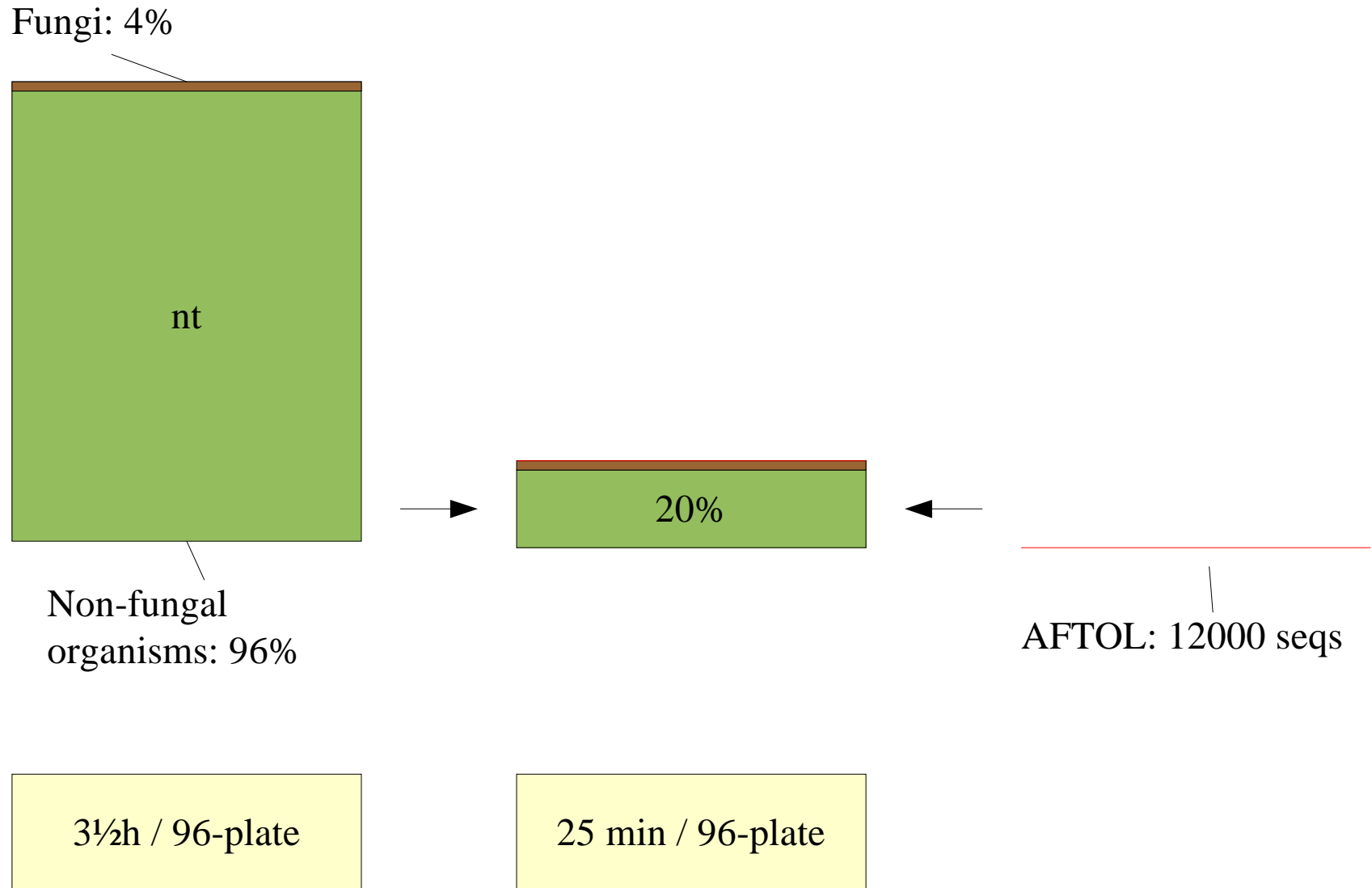
Phred & Phrap



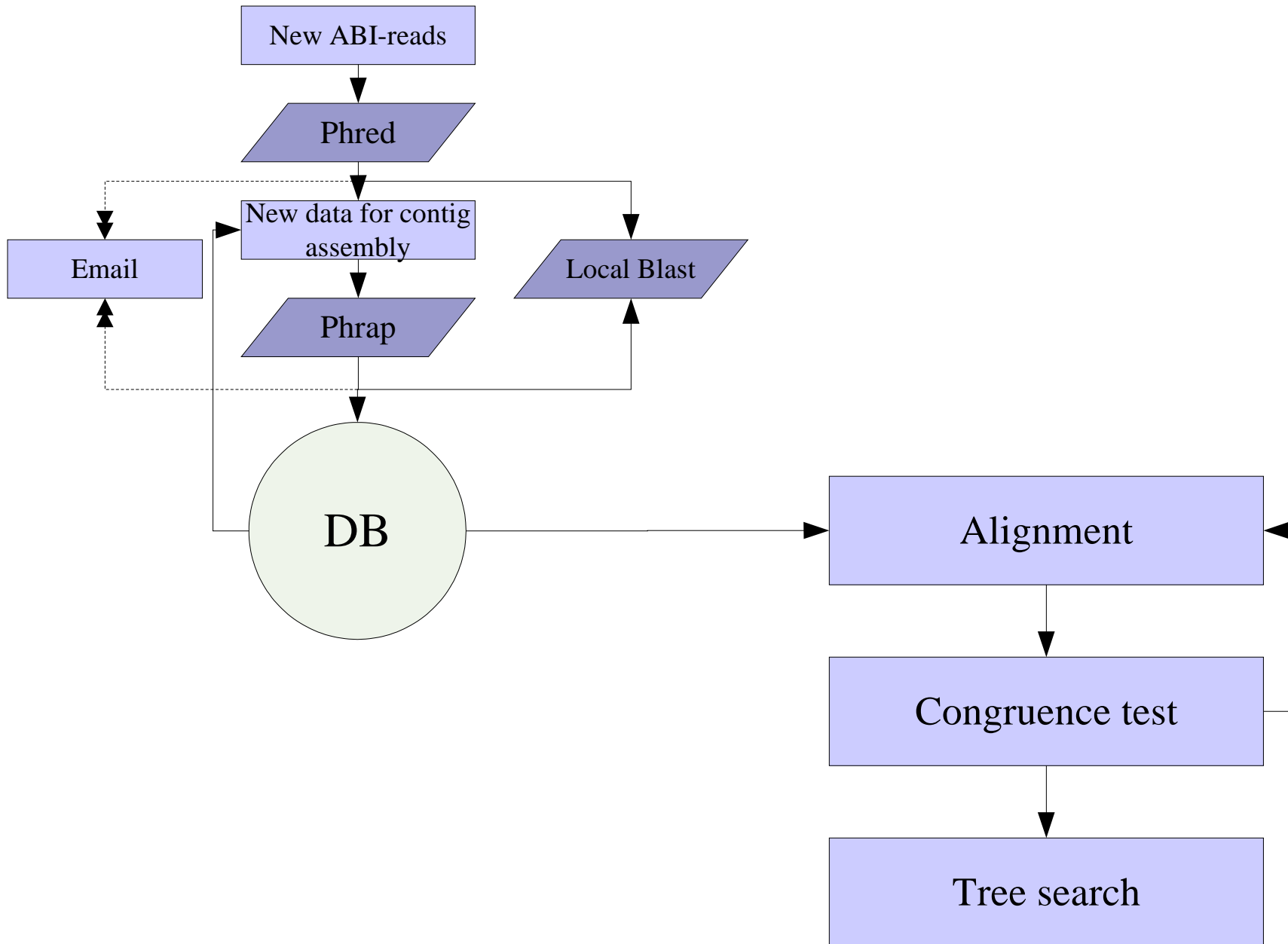
MOA: Overview



Local BLAST: custom database



Phylogenetic Analysis



Alignment

```
aTrich_hirs CTTAGGTATCGGGCGATGTTAATTTTAT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
aTrype_unkn CTTAGGGATCGGGCGATGTTATCTTTTT--ATGTCGCTCTTGGGCTCT-----GGA-----GGGGAGTATGGT
Auric_auri CTTAGGGATCGGGCGACCTCTTTTTT--ATGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Aurip_aure CTTAGGGATCGGGCGGTCTCAATTAT--ATATGTCGATCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Auris_vulg CTTAGGGATCGGGCGACCTCAATTTAA--TTTGTGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Auxar_zuff CTTAGGGATCGGGCGGGCAACTTTTAA--TATGTCGCTCTTGGGTTCTCGATCGGCTACGAGCGGACTAGCGGCGGCATCGAGCAGGGA-----GGGGAGTATGGT
aVerpa_coni CTTAGGGATCGGGCGATGCTTAATAGAT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
aXanth_cons CTTAGGGATCGGGCGGTGTTATTATTTT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
aXylar_acut CTTAGGGATCGGGCGATGTTATTTTTT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
aXylar_hypo CTTAGGGATCGGGCGATGTTATTTTTT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Backu_circ CTTAAGGATCGGGCCTGTTTATT----ATGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Backu_cten CTTAGGGATCGGGCTTGTTTATT----ATGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
BAEPLAx CTTAGGTATCGGGCGGTGTTATCATTTT--GTGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Banke_fuli CTTAGGGATCGGGCGAACTCAATTCTA--TGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Basid_hapt CTTAGGGATCGGGCAAT--GT----TATGTGCCGCTCTTAGGTTCT-----GGAAGGGCAGGATGTCGTAGGCTGGGGAGTATGGT
Basid_rana CTTAGGGATCGGGCAAT--GT----TATGTGCTGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Benja_poit CTTAGGGATCGGGCTTGTTTATT--ATGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Bimur_nova CTTAGGGATCGGGCGGTGTTTCTATTG--TGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
Blake_tris CTTAGGGATCGGGCTTGTTTATT--ATGTCGCTCTTGGGTTCT-----GGA-----GGGGAGTATGGT
```

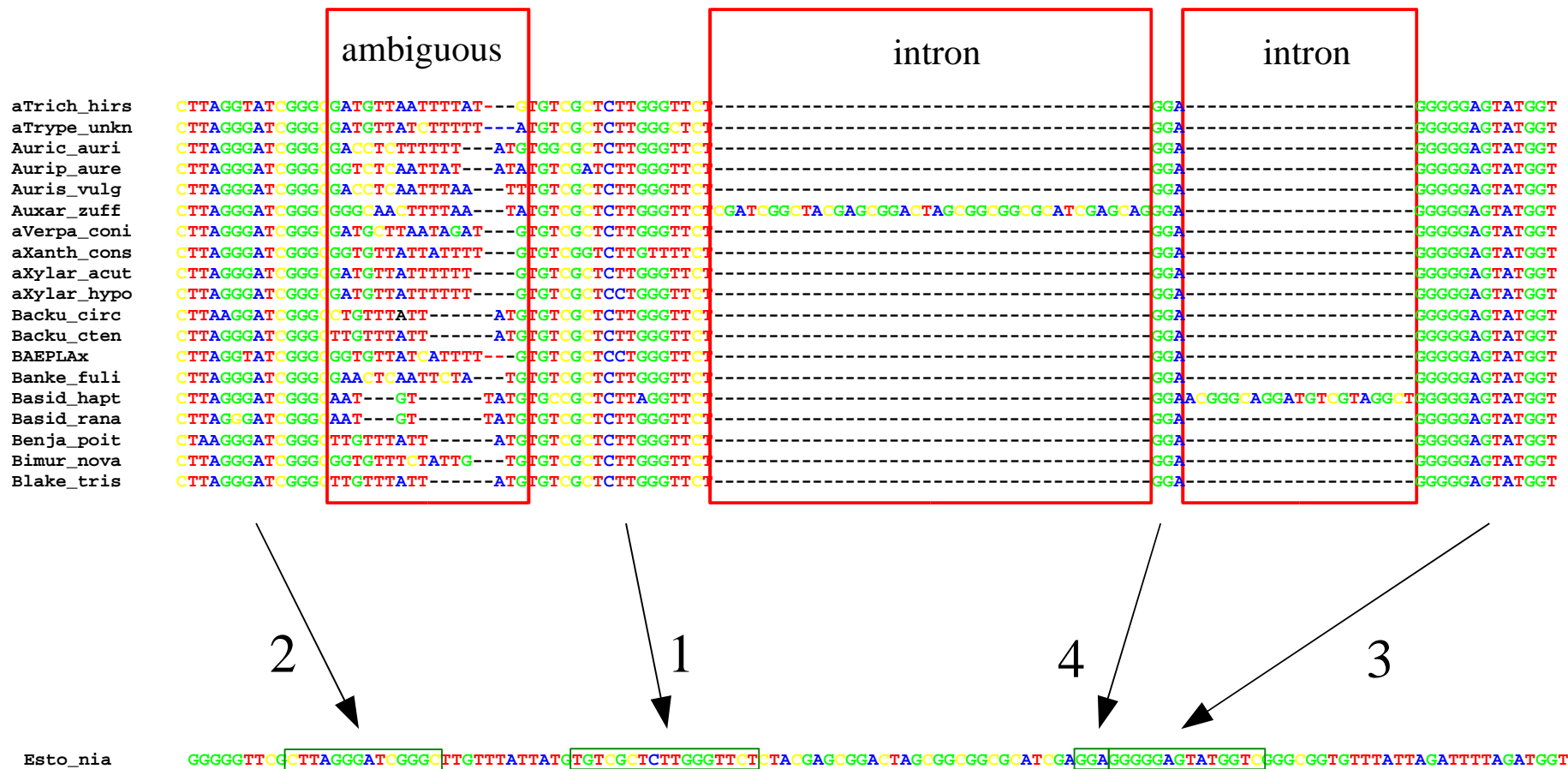
```
Esto_nia GGGGTTCTGCTTAGGGATCGGGCTTGTTTATTATGTCGCTCTTGGGTTCTCTACGAGCGGACTAGCGGCGGCATCGAGGAGGGGAGTATGGTCGGGCGGTGTTTATTAGATTTAGATGGT
```

Alignment

| | ambiguous | intron | intron |
|-------------|---|--|--------------------------------------|
| aTrich_hirs | CTTAGGTATCGGGGATGTTAATTTTAT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| aTrype_unkn | CTTAGGGATCGGGGATGTTATCTTTTT--ATGTCGCTCTTGGGCTCT | ----- | GGAT-----GGGGAGTATGGT |
| Auric_auri | CTTAGGGATCGGGGACCTCTTTTTTT--ATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Aurip_aure | CTTAGGGATCGGGGGTCTCAATTAT--ATATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Auris_vulg | CTTAGGGATCGGGGACCTCAATTTAA--TTGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Auxar_zuff | CTTAGGGATCGGGGGGCAACTTTTAA--TATGTCGCTCTTGGGTTCT | CGATCGGCTACGAGCGGACTAGCGGCGGCATCGAGCAG | GGAT-----GGGGAGTATGGT |
| aVerpa_coni | CTTAGGGATCGGGGATGCTTAATAGAT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| aXanth_cons | CTTAGGGATCGGGGGTGTTATTATTTT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| aXylar_acut | CTTAGGGATCGGGGATGTTATTTTTT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| aXylar_hypo | CTTAGGGATCGGGGATGTTATTTTTT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Backu_circ | CTTAAGGATCGGGGCTGTTTATT----ATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Backu_cten | CTTAGGGATCGGGGTTGTTTATT----ATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| BAEPLAx | CTTAGGATCGGGGGTGTTATCATTTT--GTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Banke_fuli | CTTAGGGATCGGGGAACTCAATTCTA--TGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Basid_hapt | CTTAGGGATCGGGGAAT--GT----TATGTCGCTCTTGGGTTCT | ----- | GGATCGGGCAGGATGTCGTAGGCTGGGGAGTATGGT |
| Basid_rana | CTTAGGGATCGGGGAAT--GT----TATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Benja_poit | CTTAGGGATCGGGGTTGTTTATT----ATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Bimur_nova | CTTAGGGATCGGGGGTGTTTCTATTG--TGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |
| Blake_tris | CTTAGGGATCGGGGTTGTTTATT----ATGTCGCTCTTGGGTTCT | ----- | GGAT-----GGGGAGTATGGT |

Esto_nia GGGGTTCTCGCTTAGGATCGGGGTTGTTTATTATGTCGCTCTTGGGTTCTCTACGAGCGGACTAGCGGCGGCATCGAGGAGGGGAGTATGGTCTGGGCGGTGTTTATTAGATTTAGATGGT

Alignment



The AFTOL project uses only free software and says thanks to:



Linux (OS)



Python (programming language)



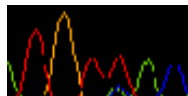
Biopython (bioinformatics modules)



PostgreSQL (SQL database)



Apache (webserver)



Phred and Phrap (Sequence tools)

- F. Lutzoni / K. Pryer labs for their support
- all AFTOL collaborators for their ideas and suggestions
- Hunter Matthews, Bill Rankin, John Pormann, Sean Dilda for computer support



This work is funded by NSF grant DEB-0228668 to Francois Lutzoni and Rytas Vilgalys.