THE BIOVEL PROJECT:
ROBUST PHYLOGENETIC WORKFLOWS RUNNING ON THE GRID

www.biovel.eu

Bachir Balech (IBBE-CNR)

The Biovel Project

BioVeL is a virtual e-laboratory that supports research on Biodiversity issues using large amounts of data from cross-disciplinary sources. It is a consortium of 15 partners from 9 countries, as well as an outer circle of 'Friends of BioVeL'

➢ Access a worldwide network of expert scientists

➢ Sharing knowledge on Biodiversity research

Biodiversity Issues

o Species identification, discovery and distributions

o The changing nature of ecosystems altering organismal composition

o The increased risks of species extinction

Decision making in biodiversity management at multiple scales (genomic, organismal, habitat, ecosystem, landscape, etc…)
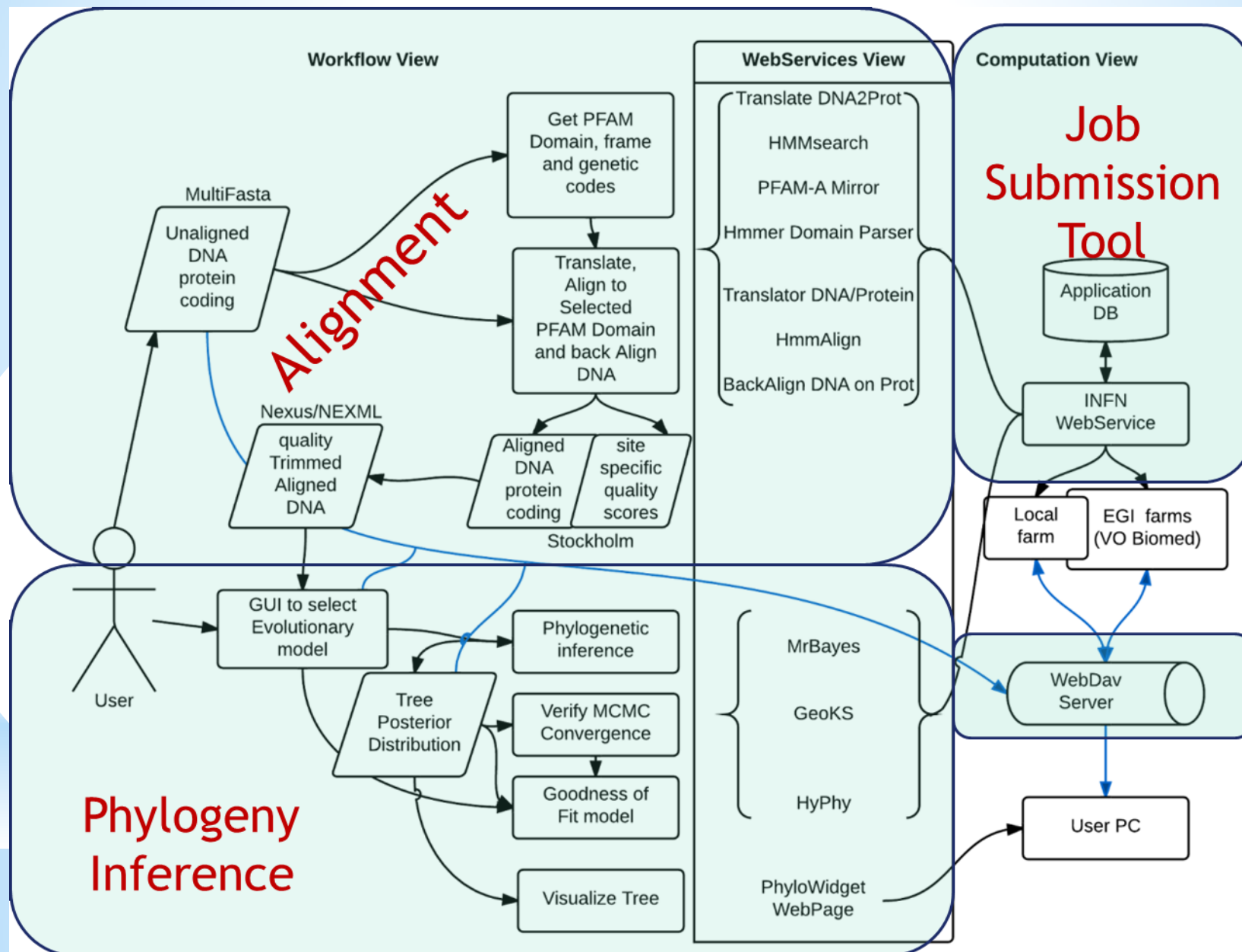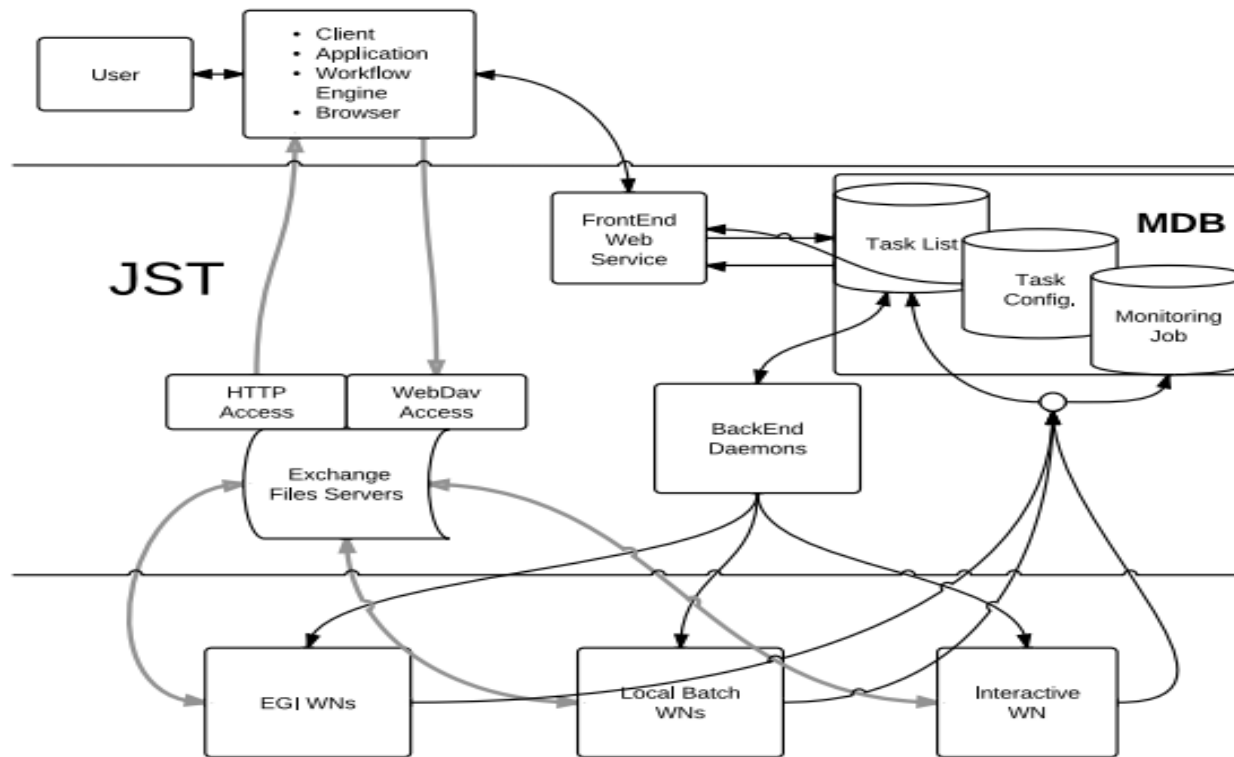
Biodiversity Solutions

✓ **Services:** data processing techniques. Each technique is available as a single executable application which can be used either alone or within a workflow builder environment (e.g. Taverna)

✓ **Workflows:** examples of services use that can be modified

---

Services and Workflows for Biodiversity Analysis:

- Taxonomy

- **Phylogenetics**

- Metagenomics

- Ecological Niche Modeling

- Ecosystem Functioning and Valuation

- Geospatial Visualization

Sharing →

BiodiversityCatalogue
"The Biodiversity Sciences Web Services Registry"

Services

myexperiment

Workflows

Example of Phylogenetic Services

Job Sumbission Tool: JST

Frontend:
- Username
- Task status
- Dependencies of each task
- Priority
- Job provenance
- Task description
- Number of failures
- Date and time of execution
- Infrastructure information (grid, local farm, interactive server)

Backend:
- Task submission at a given rate
- Stops jobs submission when no more unassigned tasks are found in the TaskList

Multiple Sequence Alignment Workflow

Multiple Sequence Alignment Workflow

File upload

Translation

HMM search

Pfam profile selection

HMM align & Back-align
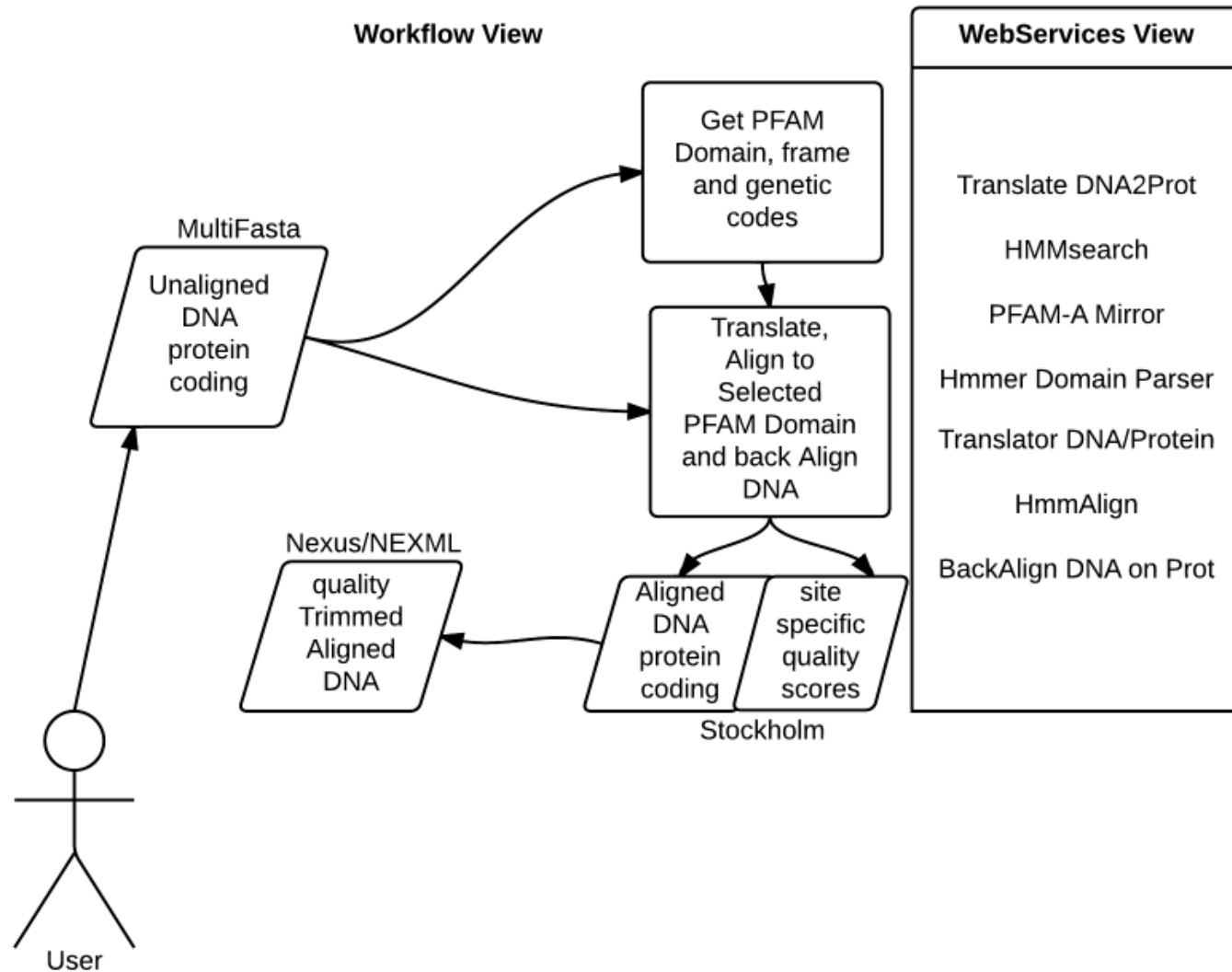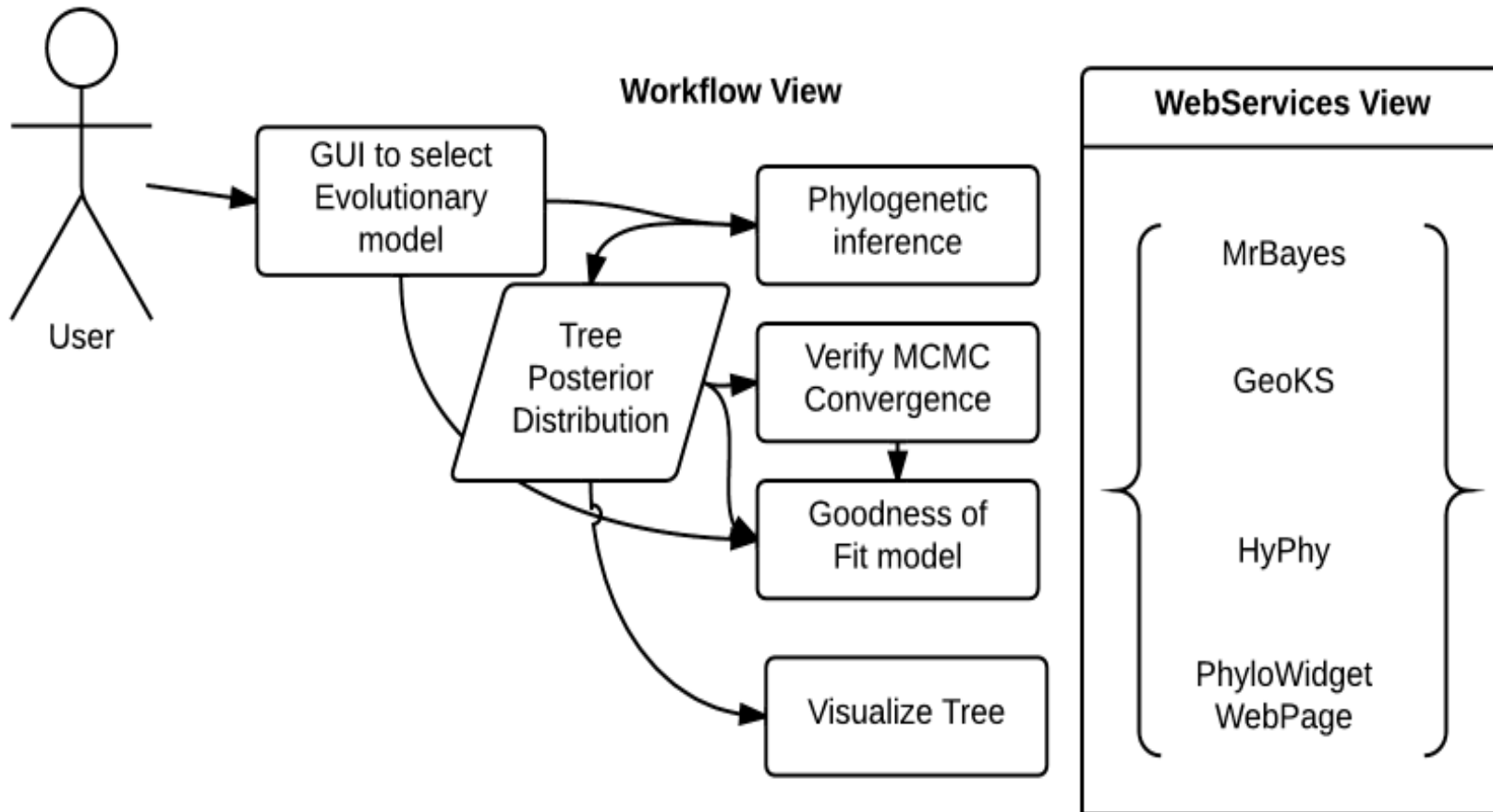
Multiple Alignment of **DNA coding**

Higher alignment precision given by:

- HMM search assigning a per site quality score (posterior probability)
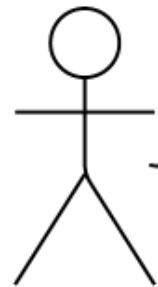- Back-align (amino acid -> DNA)

In progress:
Multiple Domain Coding sequences Alignment

Example Phylogenetic Inference Workflow

Example Phylogenetic Inference Workflow
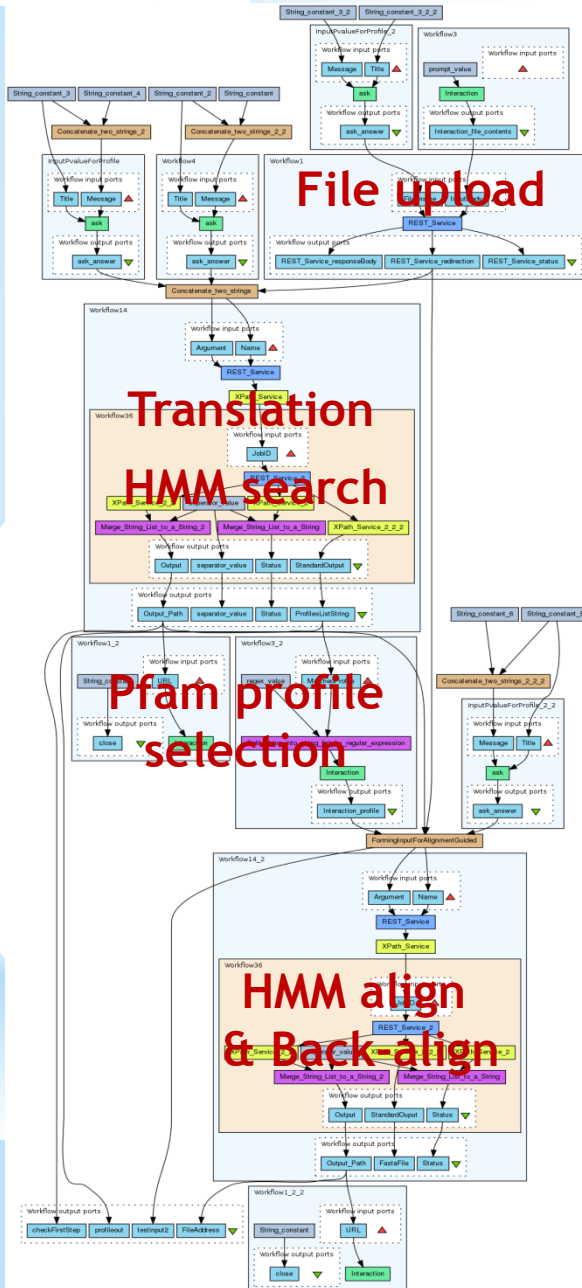
Workflow View

GUI to select Evolutionary model → Phylogenetic inference

WebServices View

MrBayes

MrBayes partitioned models

Partition 1

Partition name: Partition_1

Start: 1 End: 400 Lag: 1 [-] [+]

Evolutionary model

Model for partition 1

Transition matrix: GTR

Rates variability: gamma distribution

Lset:

Model for partition 2

Transition matrix: HKY

Rates variability: equal

Lset:

Link parameters across partitions

Matrix

Not Grouped

Partition 1 (GTR): ⦿

Partition 2 (HKY): ⦿

(🔴 = required, 🔵 = conditionally required)

Markovian Integration Parameters:
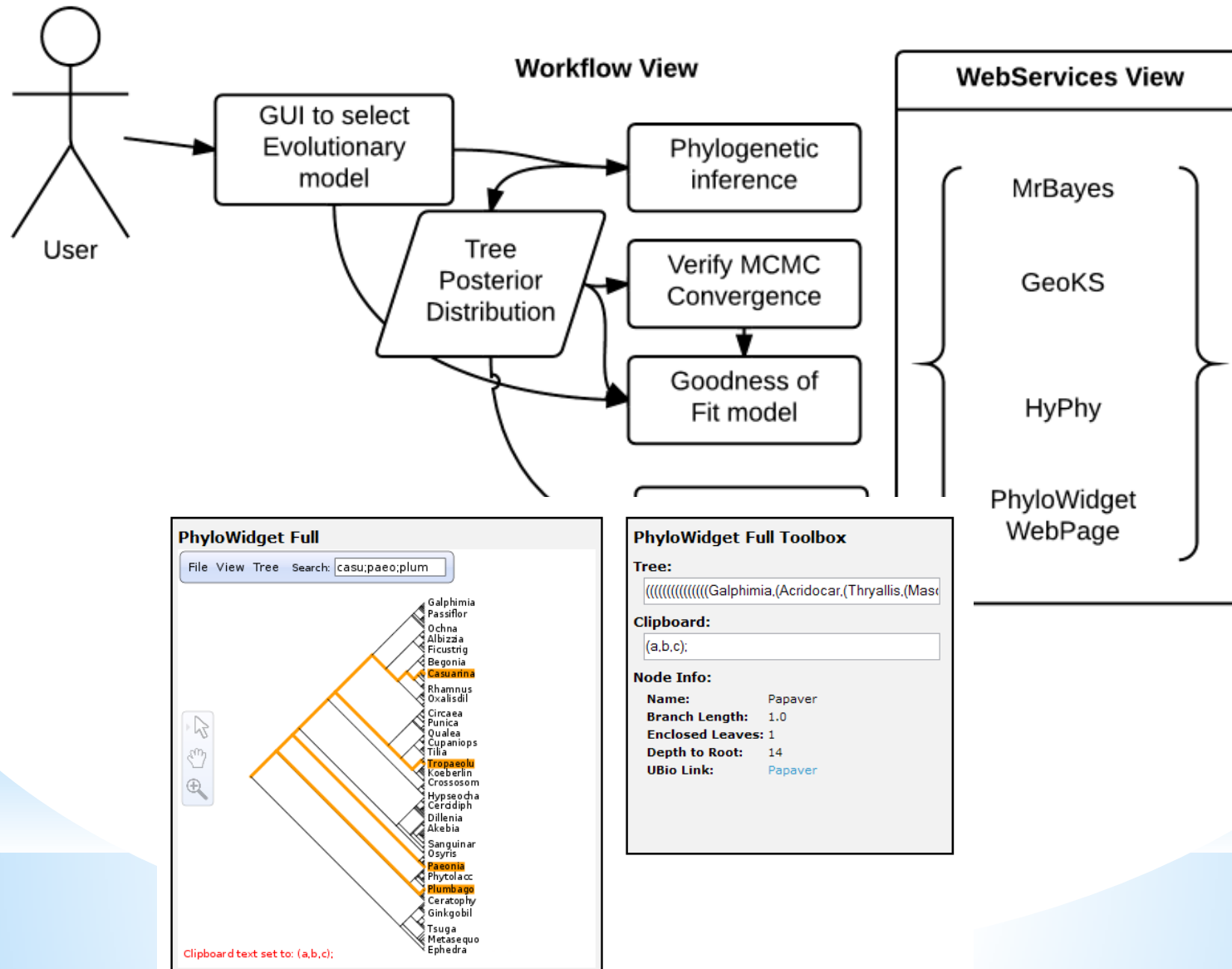
Number of generation: 1000000

Frequency of sampling: 100

Number of Chain: 4 🔴

Temperature: 0.200000

Output files prefix:

Number of Replicates: 2 🔴

MCMCP:

[Preview Formatted Nexus for MrBayes] [Reset Page]

[Send Formatted Nexus to MrBayes] [Cancel all]

Example Phylogenetic Inference Workflow

Example Phylogenetic Inference Workflow in Taverna

MrBayes Web Interface

Bayesian Phylogeny Computation & Output Retrieval

GeoKS Execution

Consensus Tree Calculation

Tree Visualization

Peculiarity:

- Partitioned models
- Convergence calculation
- Short Computation time on the Grid (even for long jobs)

Other available Phylogenetic Services:
- Maximum Likelihood (RaxML)
- Phylogenetic Diversity (Phylocom)

Acknowlegments

Bioinformatic Scientists
Prof. Graziano Pesole
Dr. Saverio Vicario

ICT specialists
Dr. Giacinto DONVITO
Dr. Pasquale NOTARANGELO