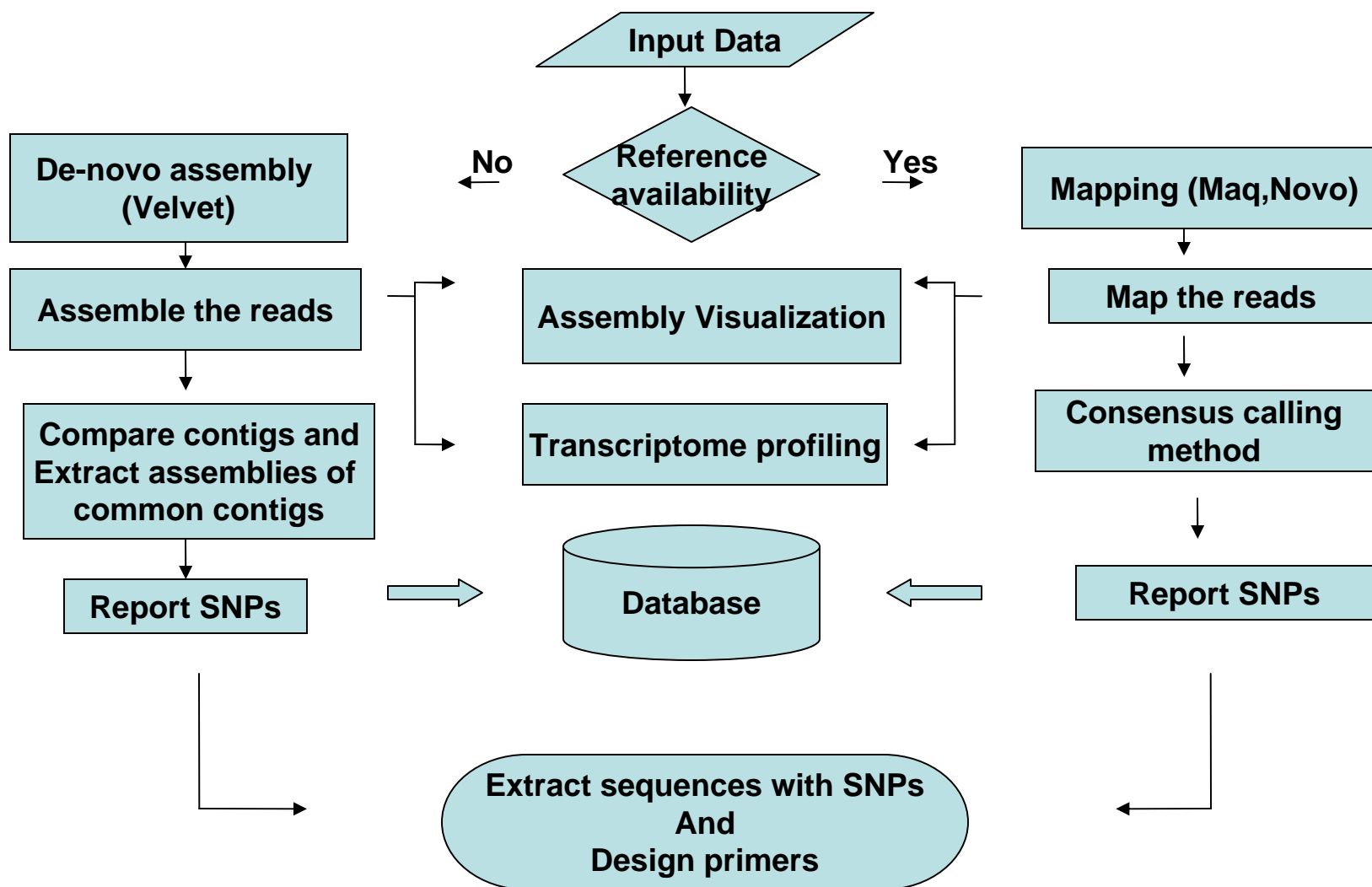# NGS Data Analysis Pipeline @ ICRISAT

Pradeep R
Bioinformatics

Control Flowchart

User friendly GUI interface.
Linux platform
Data communication with java

Computational resources
- Parallel computing on HPC
- Compute cluster

Input format compatibility
- FASTA
- FASTA and QUAL
- Solexa & Sanger (FASTQ)

Mapping/Assembly
- Up to two genotypes.

On fly documentation for software
- Maq(v0.6.8)
- Novo (v2.05.02)

Reference database
– Publicly available assemblies (Genomic and transcriptome) of major crops

Session specific job processing
- Avoid file over writing

Accessibility of intermediate output files
- FastQ
- Binary files (*.bfa,*.bfq)
- Map files (*.map)
- Consensus Sequence file
- Pileup file

Acknowledgement through SMTP server
- Posting results through Email

Interactive graphical visualization with Gbrowse
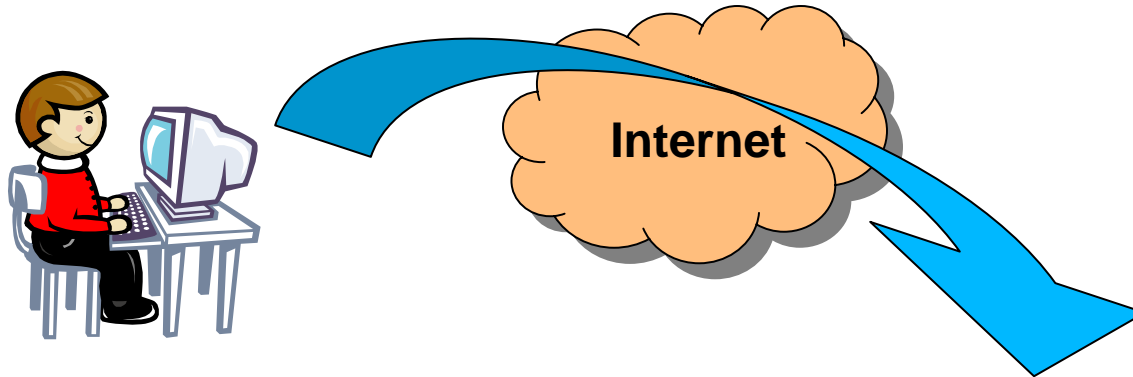- Reference
- Assembly (mapped reads)

SNPs
- variants between assembly and reference
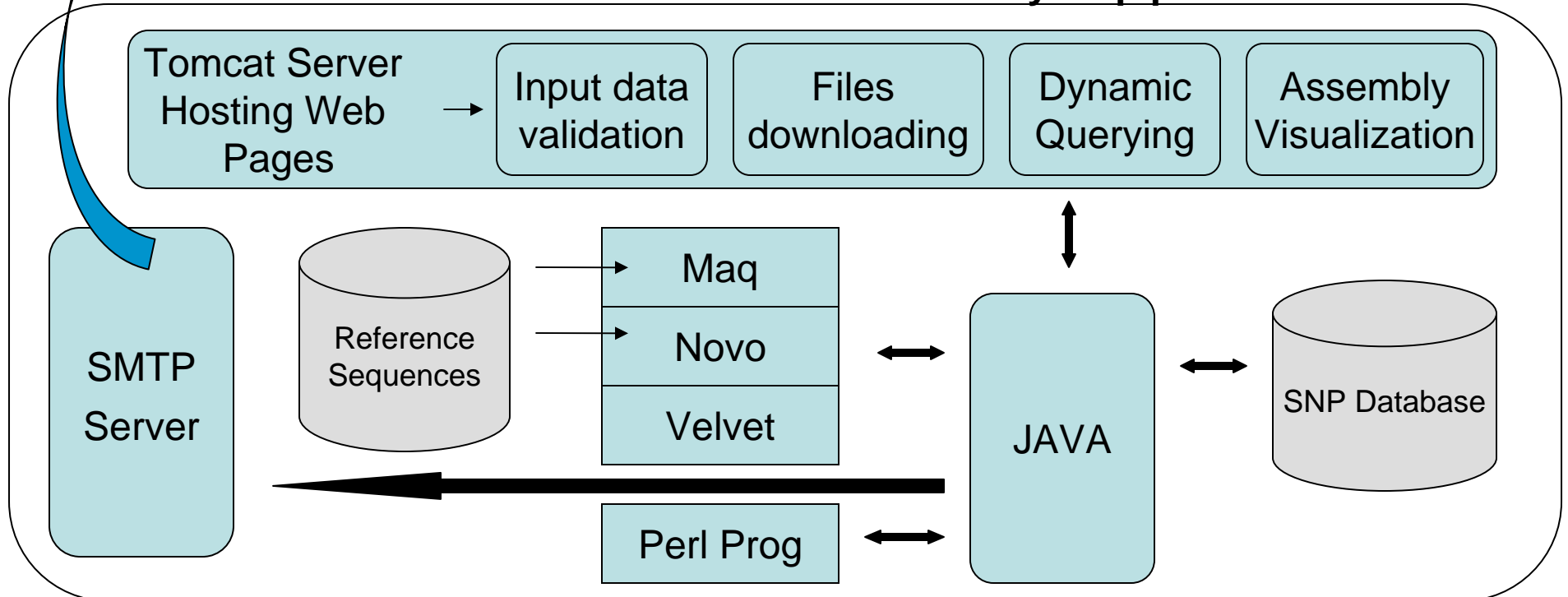- variants between assemblies of two genotypes

Analysis of Transcriptome profile
- Reads Per Kb (of reference) Per Million mapped reads (RPKM)

Architecture

ICRISAT
Science with a human face

Internet

NGS Data Analysis pipeline at ICRISAT

Tomcat Server Hosting Web Pages → Input data validation | Files downloading | Dynamic Querying | Assembly Visualization

SMTP Server

Reference Sequences

Maq
Novo
Velvet

JAVA

SNP Database

Perl Prog

Acknowledgements

- Dr. Rajeev K. Varshney

- Dr. Jayashree B

- Dr. Vivek Thakur

- Mr. A. Bhanu Prakash

ICRASAT
Science with a human face

Drop

Suggestion and comments

To

r.pradeep@cgiar.org

a.bhanuprakash@cgiar.org