

WORKSHOP

Marky Coco



marky coco 2

Comprehensive Tool
for Mercury Genes
Detection

Marky Coco²

-
-
- Automatic validation of hgcA and hgcB
- Automatic validation of merA and merB
- Generation of merA and merB for MAGs, SAGs, and isolates
- Taxonomic assignment of the genes
- Automatic gene coverage normalization with different normalization techniques
-
-
-

**Marky
Coco** **2**



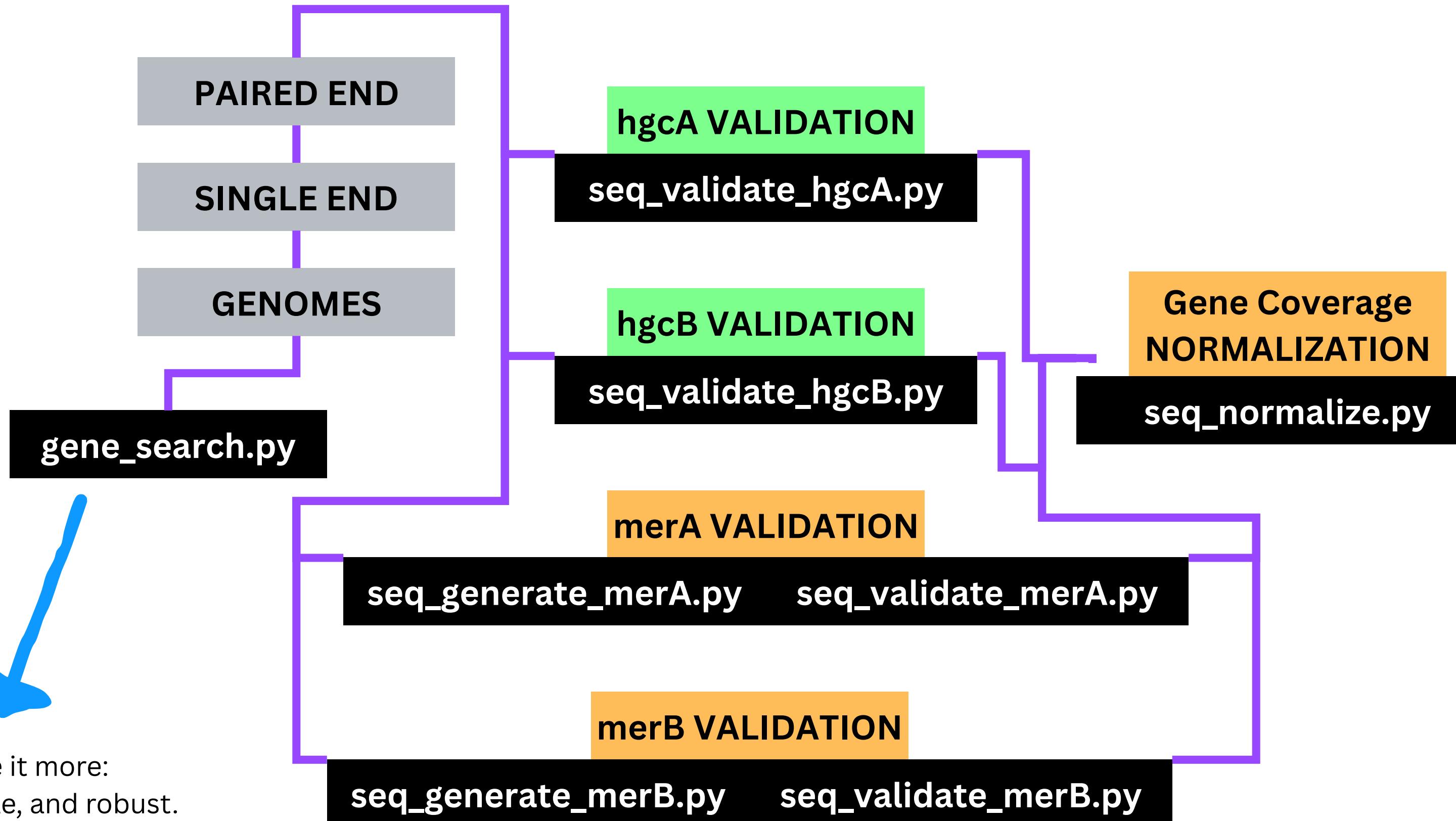
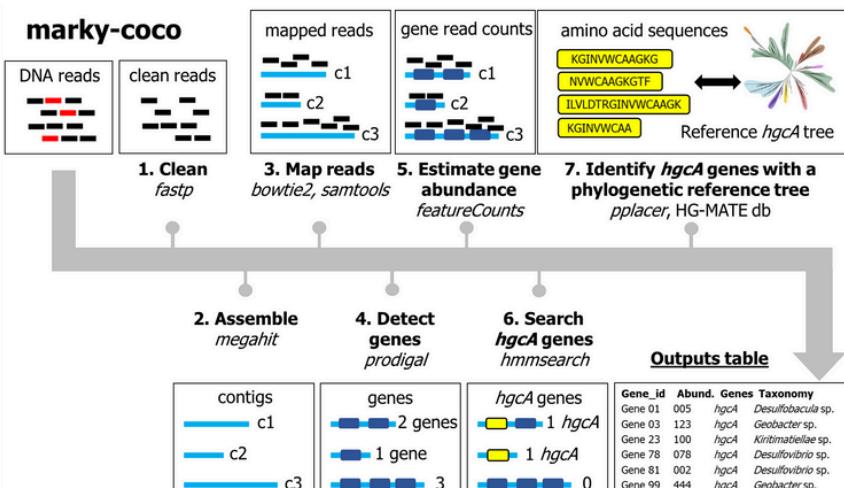
**Marky
Coco** **2**
alpha

-
-
- **Automatic validation of hgcA and hgcB**
- **Automatic validation of merA and merB**
- **Generation of merA and merB for MAGs, SAGs, and isolates**
- Taxonomic assignment of the genes
- Automatic gene coverage normalization with different normalization techniques
-
-
-

Marky Coco 2

alpha

Marky Coco



Conversion to Python

Python will make it more:
maintainable, readable, and robust.

It would eliminate:
temporary files

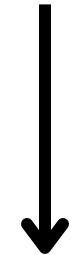
long chains of awk/sed/grep.
repetition of large blocks of code



merA and merB validation

GENOMES

- Generation of merA and merB:
for each BIN it run **hmmsearch** and **esl-reformat** to produce one table
and one FASTA file for merA and merB



PAIRED END

- Validation of merA and merB

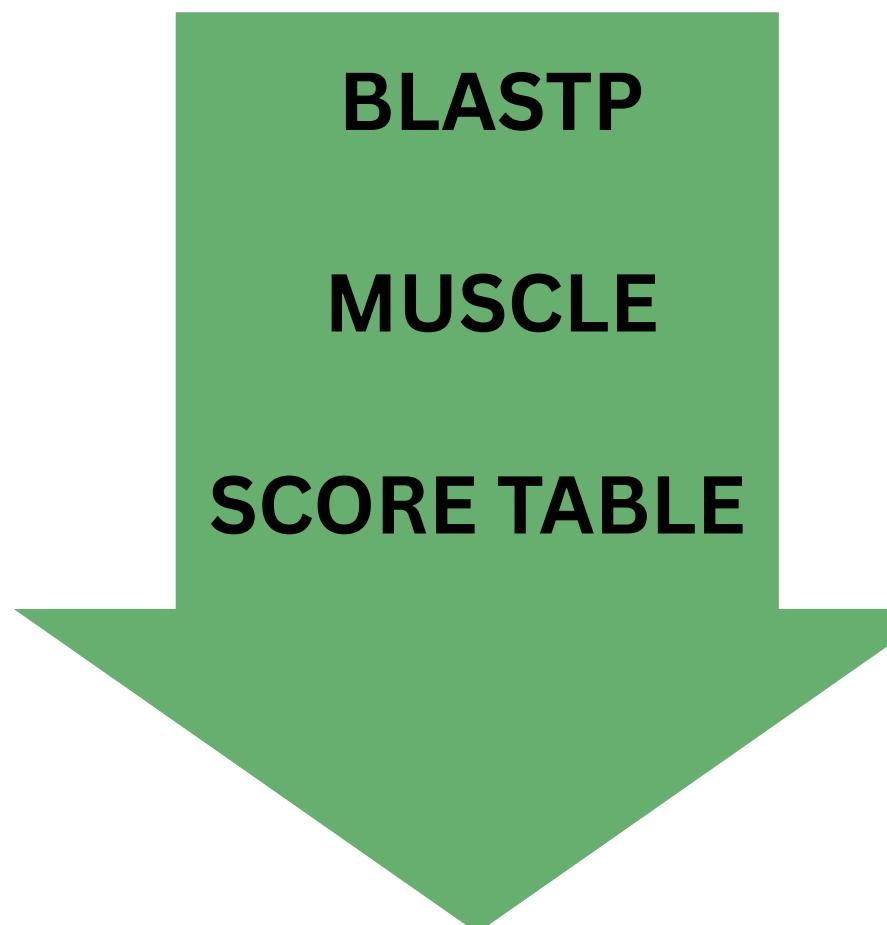
SINGLE END

merA validation

*merA: *Bacillus* sp. RC607

merB: *Escherichia coli* plasmid R831b

multi-step validation



Number of Sequences

Tabular file

seq_id | length | read | cov | contig | sequence | gene_type

high-confidence
merA sequences

The score table

The following MerA sequence signatures have been experimentally shown to be essential for MerA activity (Barkay et al., 2003):

- the conserved cysteine pair at positions 207 and 212
- the vicinal cysteine pair at the carboxy terminus (positions 628 and 629),
- tyrosine at position 264 (Rennex et al., 1993)

C207 C212 C628 C629 Y264

1 1 1 1 1

0 to 5 points

Aligned MerB homologs were examined for the presence of the following sequence signatures:

- cysteine at position 96 (numbering in reference to MerB from R831b)
- aspartic acid at position 99,
- and cysteines at position 159 and 117

Following Pitts and Summers, 2002 and Lafrance-Vanasse et al., 2009.

D99 C159 C117 C96

1 1 1 1

0 to 4 points

The score table

DIFFERENT THRESHOLD



DIFFERENT NUMBER OF SEQUENCES

The following MerA sequence signatures have been experimentally shown to be essential for MerA activity (Barkay et al., 2003):

- the conserved cysteine pair at positions 207 and 212
- the vicinal cysteine pair at the carboxy terminus (positions 628 and 629),
- tyrosine at position 264 (Rennex et al., 1993)

C207	C212	C628	C629	Y264
1	1	1	1	1

0 to 5 points

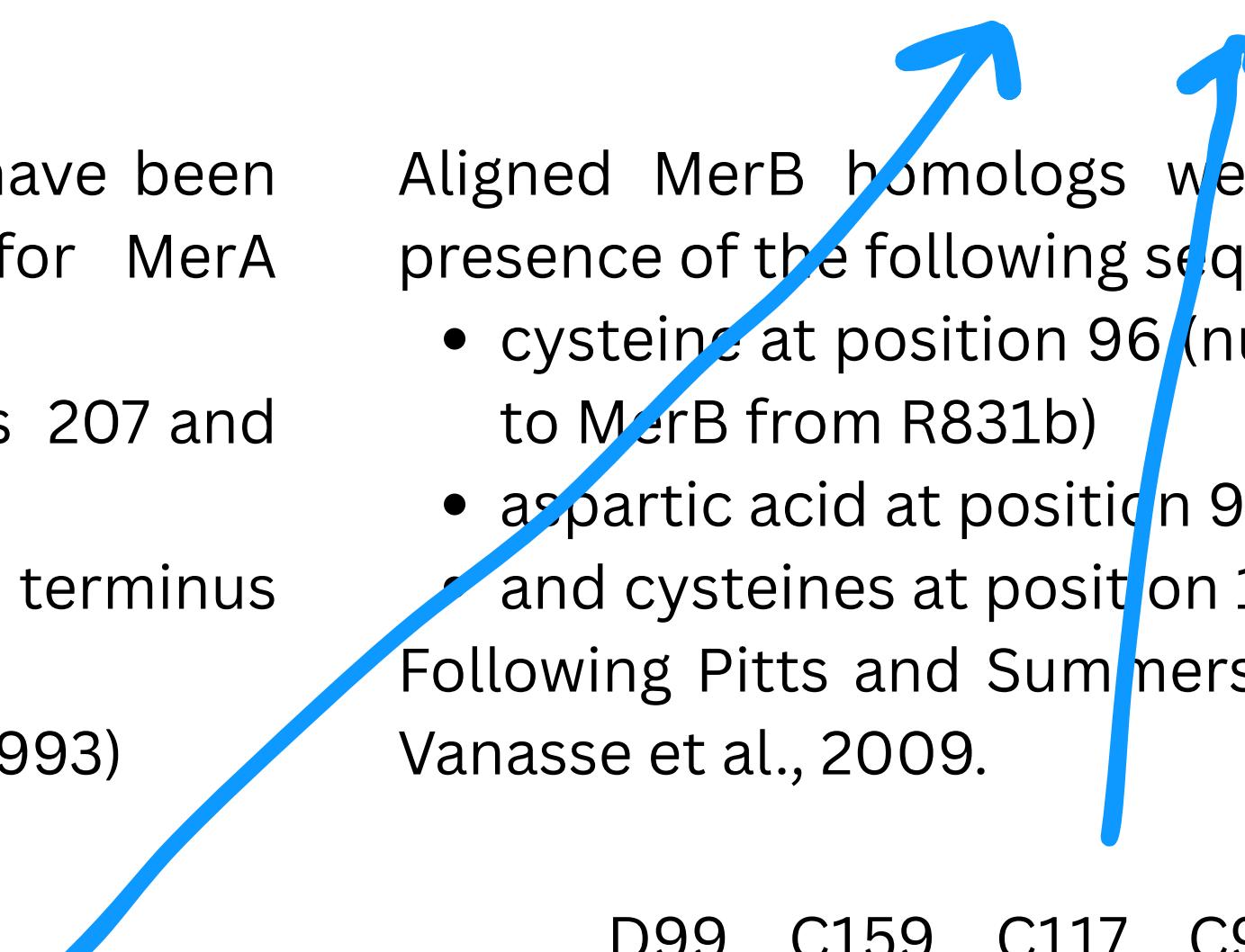
Aligned MerB homologs were examined for the presence of the following sequence signatures:

- cysteine at position 96 (numbering in reference to MerB from R831b)
- aspartic acid at position 99,
- and cysteines at position 159 and 117

Following Pitts and Summers, 2002 and Lafrance-Vanasse et al., 2009.

D99	C159	C117	C96
1	1	1	1

0 to 4 points





Hands
On

Questions





Thank You
For Your Attention