

De Novo Genome Analysis

Ketil Malde

Institute of Marine Research

October 2, 2012

Overview

De Novo

Ketil Malde

Introduction

Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis

Introduction

Assembly

Gene prediction

Assembly evaluation

Annotation

Analysis

Processing a new genome

Introduction

Assembly

Gene prediction

Assembly evaluation

Annotation

Analysis

1. Construct draft genome assembly
2. Map known sequences
(ESTs, RNAseq, proteins)
3. Build/train stochastic gene models
4. Predict genes and transcripts
5. Annotate transcripts with function
6. Do awesome biology stuff - yay!

The pipeline

De Novo

Ketil Malde

Introduction

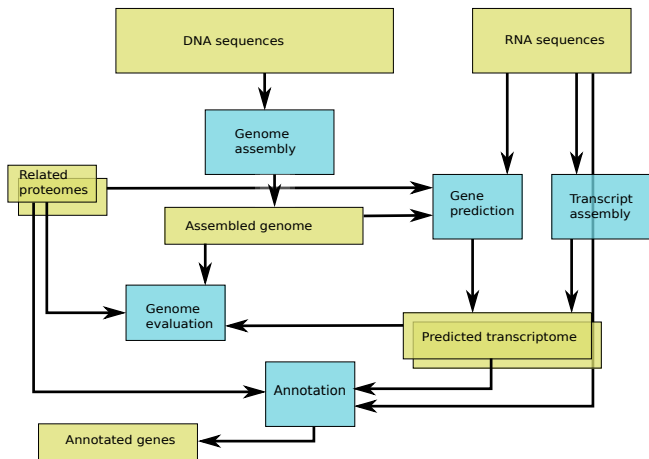
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Assembly

De Novo

Ketil Malde

Introduction

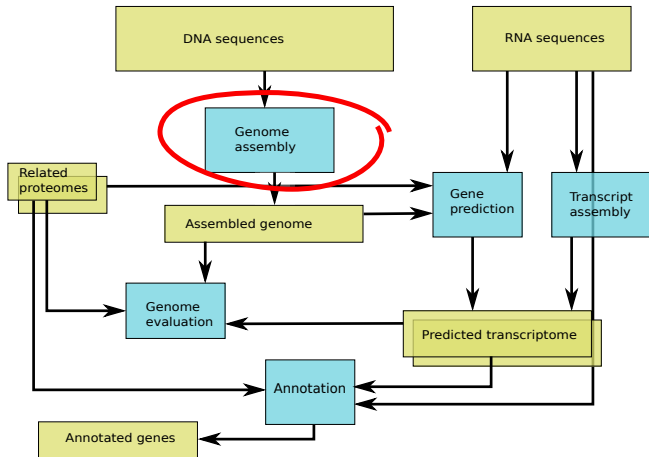
Assembly

Gene prediction

Assembly evaluation

Annotation

Analysis



Assembly

De Novo

Ketil Malde

Introduction

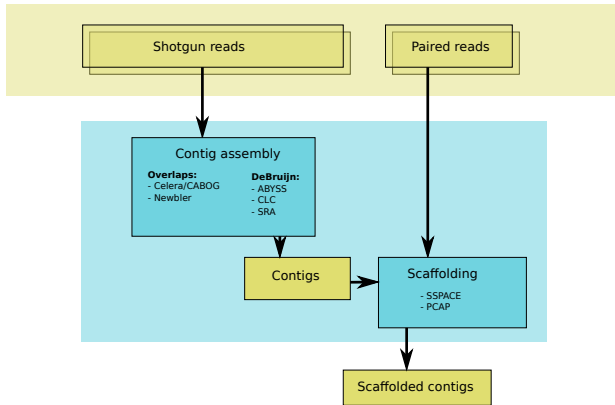
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Assembly

De Novo

Ketil Malde

Introduction

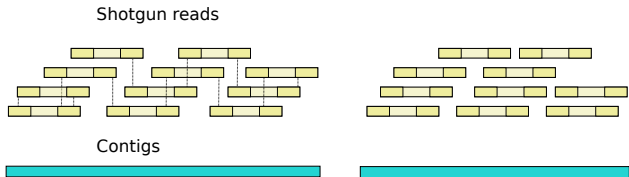
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Assembly

De Novo

Ketil Malde

Introduction

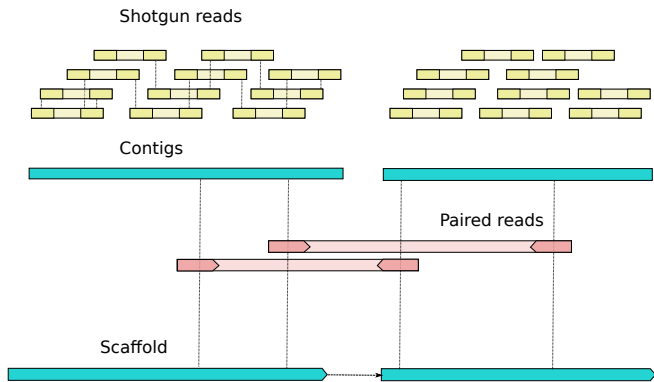
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Gene prediction

De Novo

Ketil Malde

Introduction

Assembly

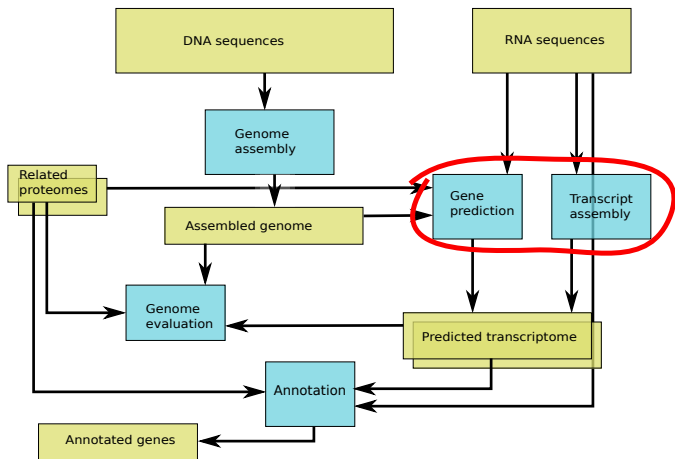
Gene prediction

Assembly

evaluation

Annotation

Analysis



Gene prediction

De Novo

Ketil Malde

Introduction

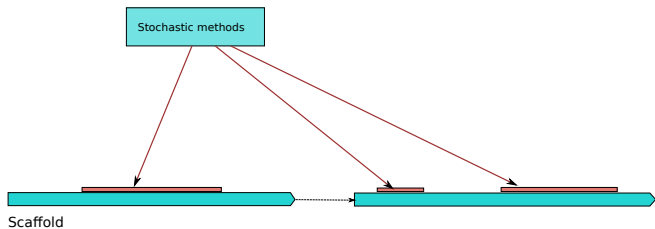
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Gene prediction

De Novo

Ketil Malde

Introduction

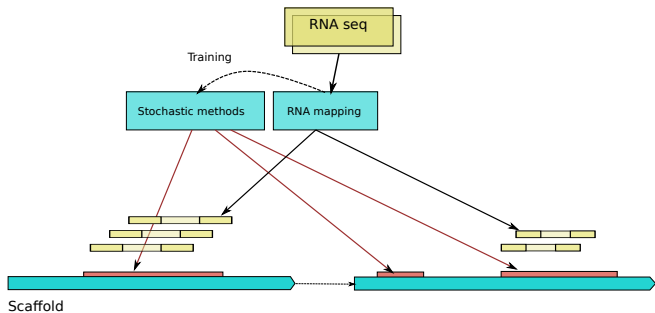
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Gene prediction

De Novo

Ketil Malde

Introduction

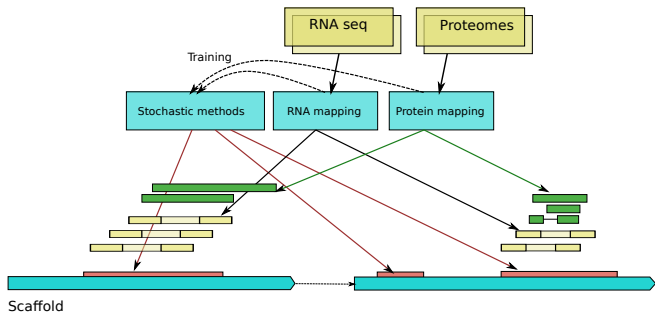
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Gene prediction

De Novo

Ketil Malde

Introduction

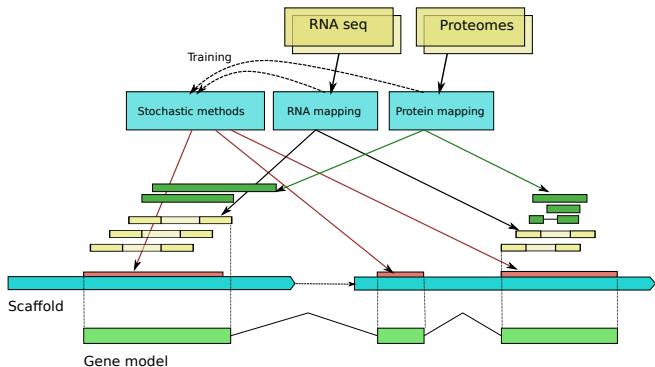
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Assembly evaluation

De Novo

Ketil Malde

Introduction

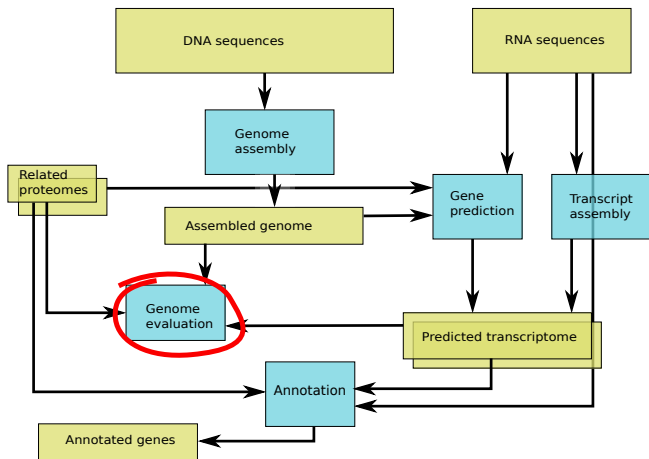
Assembly

Gene prediction

Assembly evaluation

Annotation

Analysis



Assembly evaluation

- ▶ Accuracy
 - ▶ read mapping scores
- ▶ Completeness
 - ▶ mapped reads
 - ▶ mapped proteins
- ▶ Fragmentation
 - ▶ N50 etc.
 - ▶ mapped pairs
- ▶ Redundancy
 - ▶ total scaffold size
 - ▶ multiply mapped reads
 - ▶ multiply mapped proteins

Annotation

De Novo

Ketil Malde

Introduction

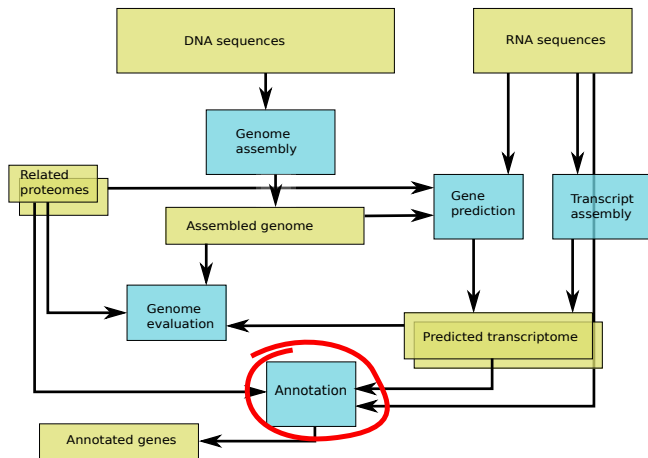
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Annotation

De Novo

Ketil Malde

Introduction

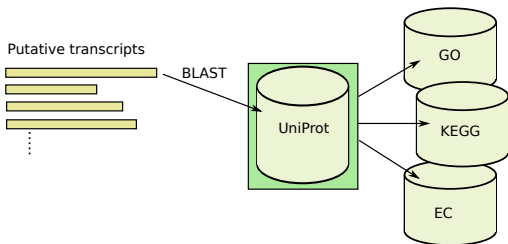
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Transitive alignments

De Novo

Ketil Malde

Introduction

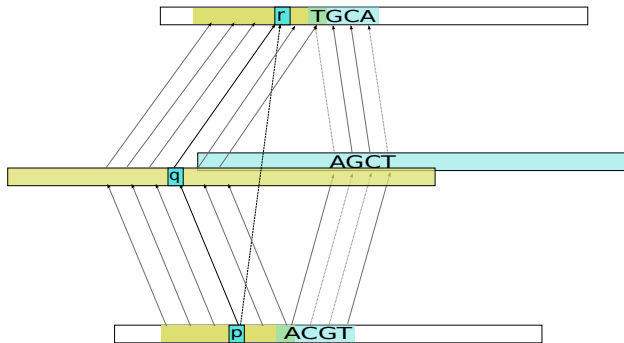
Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis



Awesome biology stuff!

De Novo

Ketil Malde

Introduction

Assembly

Gene prediction

Assembly
evaluation

Annotation

Analysis

