Recognition of Polyadenylation Sites from Genomic Arabidopsis Sequences

Chuan Hock Koh
Limsoon Wong
Human Genome Project

• Completed in 2003
• What's next?
  – Genomes to Life (GTL) Project
    • Understanding the detail mechanism of the cell
  – ENCODE Project
    • Annotating the whole Genome
Research Project

• Phase 1
  – Develop a computational method to analyze (annotate) DNA sequences

• Phase 2
  – Make this method easy to use, especially for biologists with no or little computer expertise
• Develop a 4 step methodology
  1. Feature Generation
  2. Feature Selection
  3. Feature Integration
  4. Cascade Classification

• 18th International Conference on Genome Informatics
1. Feature Generation
   • Thickness of Eyebrows, Hair Length, Hair Color, Jaw Line, Eye Color etc

1. Feature Selection
   • Reject Hair Length and Hair Color

1. Feature Integration
   • Use all selected features to judge
1. Cascade Classifier
   • Looking at its surroundings
Phase 2

- To enable the Method to be deployed with ease
  - Developed a software
Conclusion

• Personal Genome Project
  – Individual genome for < USD 1000

• Develop more efficient ways to store, retrieve and 'crunch' data