

# Contents

## Part I Introduction to Molecular Biology and Bioinformatics Basics in Next-Generation Sequencing

<b>1 Basics of Molecular Biology for Next-Generation Sequencing.....</b>	<b>3</b>
1.1 Molecular Biology .....	3
1.1.1 Deoxyribonucleic Acid .....	5
1.1.2 Ribonucleic Acid.....	6
1.1.3 Proteins.....	8
1.2 The Central Dogma of Molecular Biology .....	9
1.2.1 Transcription .....	9
1.2.2 Translation.....	10
1.3 Genetic Information Sources Targeted by Sequencing .....	10
1.3.1 The Genome .....	11
1.3.2 The Transcriptome .....	11
1.3.3 The Exome .....	12
1.3.4 The Metagenomes .....	12
References .....	12
<b>2 Algorithms and Data Structures in Next-Generation Sequencing .....</b>	<b>15</b>
2.1 Data Structures .....	15
2.1.1 Strings .....	15
2.1.2 Suffix/Prefix Trees.....	16
2.1.3 Suffix Arrays .....	16
2.1.4 Graphs .....	17
2.1.5 Hash Tables .....	19
2.1.6 Bloom Filters.....	20
2.2 Algorithms .....	21
2.2.1 Alignment Algorithms .....	21
2.2.2 Greedy Algorithm .....	23
References .....	24

## **Part II Next-generation Sequencing Methods, Platforms, Applications and Challenges**

<b>3 First- and Next-Generations Sequencing Methods .....</b>	<b>29</b>
3.1 Introduction.....	29
3.2 DNA Sequencing Methods .....	29
3.2.1 First Generation Sequencing Methods .....	30
3.2.2 Next-Generation Sequencing Methods .....	32
References .....	35
<b>4 Next-Generation Sequencing Platforms .....</b>	<b>37</b>
4.1 Introduction.....	37
4.2 Roche 454 Systems .....	37
4.3 AB SOLiD Systems .....	39
4.4 Illumina GA/HiSeq Systems.....	40
4.5 Compact Systems.....	42
4.5.1 Illumina MiSeq.....	43
4.5.2 Ion Torrent PGM .....	43
4.5.3 The Open-Source Sequencing System .....	43
References .....	44
<b>5 Challenges in the Next-Generation Sequencing Field .....</b>	<b>45</b>
5.1 Sanger Sequencing Versus Next-Generation Sequencing.....	45
5.2 Efficiency of Sanger Sequencing .....	46
5.3 Challenges in Next-Generation Sequencing .....	47
5.3.1 Sequencing Cost.....	47
5.3.2 Read Length .....	47
5.3.3 Error Rates and Types .....	48
References .....	49
<b>6 New Horizons in Next-Generation Sequencing .....</b>	<b>51</b>
6.1 Third-Generation Sequencing Methods .....	51
6.1.1 Heliscope Single-Molecule Sequencing .....	51
6.1.2 Single-Molecule Real-Time Sequencing .....	52
6.1.3 Nanopore Sequencing .....	52
6.2 Third-Generation Sequencing Platforms .....	53
6.2.1 HeliScope Single-Molecule Sequencer.....	53
6.2.2 PacBio RS II.....	53
6.2.3 Oxford Nanopore GridION .....	55
6.3 Sequencing Methods Under Development.....	55
6.3.1 Solution-Based Hybridization Sequencing .....	56
6.3.2 Tunneling Current DNA Sequencing .....	56
6.3.3 Microscopy-Based DNA Sequencing .....	56
6.3.4 Mass Spectrometry-Based DNA Sequencing .....	57
6.3.5 RNA Polymerase Sequencing .....	57
References .....	58

<b>7 Novel Next-Generation Sequencing Applications .....</b>	<b>61</b>
7.1 Introduction.....	61
7.2 Large-Scale Applications.....	62
7.2.1 Genome 10K Project.....	62
7.2.2 Tree of Life Sequencing Project.....	63
7.3 Cell and Cell Compartments Applications.....	63
7.3.1 Single-Cell Genome Sequencing .....	63
7.3.2 Mitochondrial Genome Sequencing.....	64
7.4 Disease-Targeted Sequencing .....	64
7.4.1 Sequencing in Cancer.....	65
7.4.2 Sequencing in Inherited Human Diseases.....	66
7.4.3 Sequencing in Complex Human Diseases.....	67
References.....	67

### **Part III Next-generation Sequence Assembly Stages, Assessments, Tools and Challenges**

<b>8 Next-Generation Sequence Assembly Overview .....</b>	<b>73</b>
8.1 Introduction to Next-Generation Sequence Assembly.....	73
8.2 Sequence Assembly Framework.....	74
8.2.1 Error Correction Phase.....	75
8.2.2 Graph Construction Phase.....	75
8.2.3 Graph Simplification Phase.....	76
8.2.4 Scaffolding Phase.....	76
8.2.5 Assembly Assessment Phase.....	76
References.....	77
<b>9 Approaches and Challenges of Next-Generation Sequence Assembly Stages .....</b>	<b>79</b>
9.1 Introduction.....	79
9.2 Error Correction Phase.....	80
9.2.1 <i>K</i> -mer-Based Error Correction.....	80
9.2.2 Suffix Tree/Array-Based Error Correction.....	81
9.2.3 Alignment-Based Error Correction .....	81
9.3 Graph Construction Phase.....	82
9.3.1 Overlap-Based Graph Construction Approach .....	84
9.3.2 <i>K</i> -mer-Based Graph Construction Approach .....	85
9.3.3 Greedy-Based Graph Construction Approach .....	86
9.4 Graph Simplification Phase.....	86
9.5 Scaffolding Phase.....	87
9.5.1 Scaffolders.....	88
References.....	91

<b>10</b>	<b>Assessment of Next-Generation Sequence Assembly .....</b>	<b>95</b>
10.1	Introduction to Assembly Assessment.....	95
10.2	Contiguity and Consistency Measures.....	95
10.2.1	Contiguity Assessment.....	95
10.2.2	Consistency Assessment.....	96
10.3	Accuracy Measures .....	96
10.4	Assembler’s Performance Measures .....	97
10.5	Assessment Tools and Evaluation Studies for Assessing Assembly Quality .....	97
10.5.1	Evaluation Studies for Assessing Assembly Quality .....	97
10.5.2	Assembly Assessment Tools .....	98
10.6	Assessment of Transcriptome and Metagenomes Assembly Quality.....	99
	References .....	99
<b>11</b>	<b>Next-Generation Sequence Assemblers.....</b>	<b>103</b>
11.1	Introduction.....	103
11.2	Next-Generation Genome Assemblers.....	106
11.2.1	Overlap-Based Assemblers .....	106
11.2.2	<i>K</i> -Spectrum-Based Assemblers.....	110
11.2.3	Greedy-Based Assemblers .....	111
11.2.4	Hybrid Assemblers.....	112
11.3	Next-Generation Transcriptome Assemblers.....	113
	References .....	114
	<b>Concluding Remarks .....</b>	<b>117</b>

Next Generation Sequencing Technologies and  
Challenges in Sequence Assembly

El-Metwally, M.Sc, S.; Ouda, O.M.; Helmy, M.

2014, XII, 118 p. 12 illus., 1 illus. in color., Softcover

ISBN: 978-1-4939-0714-4