

INDEX

A

Aedes aegypti..... 215, 218, 228
 Agencourt AMPure XP reagent 207
 Agilent Bioanalyzer 2100 Expert Software 208
 Amazon Web Services (AWS)..... 4
 Aneuploidy 16
Anopheles arabiensis..... 67
Anopheles gambiae..... 214, 215, 218, 222, 225, 229
 Anthocyanin 19
 Arthropod..... 45, 47, 50, 53, 75
 Assembly
 10× assembly correction and scaffolding..... 38
 Canu..... 36, 37, 42
 eukaryotic 42
 genome 33, 34, 37–42
 heterozygous genomes 41
 HiRise 38
 Miniasm 35, 36
 minimum RAM 41
 PacBio 42
 practices 7, 10
 software..... 34

B

Bacterial artificial chromosome (BAC) 6
 BEDTools program..... 77, 122, 124, 132, 133
Bemisia tabaci 41
 Benchmarking Universal Single-Copy Orthologue
 (BUSCO)..... 40, 42, 60–72
 Beta-mercaptoethanol (BME)..... 28
 Bioinformatics
 biolinux 161
 DNA methylation 158, 160
 Illumina sequencing data 160
 PCR duplication 160
 quality control 162
 software programs 160
 tools 161–162
 Biopython 49, 50
 Bismark 161, 164, 167–169, 173
 Bisulfite sequencing
 bioinformatic analysis..... 158, 170
 bismark—bowtie2 161, 167–169
 bisulfite conversion errors..... 159
 CpG methylation 158

data input requirements..... 160
 DNA methylation 157, 158, 160
 epigenomic information 157
 FastQC..... 161–165
 hardware requirements 160
 issues 173
 PCR bias 160
 R—bioconductor 161
 Seqmonk..... 162
 sodium bisulfite treatment..... 159
 standard DNA sample..... 158
 statistical analysis 169, 171, 172
 subsequent PCR amplification 159
 trim galore—trimmomatic—cutadapt .. 161, 165–167

C

Caffeine 19
Callosobruchus maculatus 22
 Canu assembly..... 36
 Carnoy's solution 182, 184, 186, 188, 192
 Chelicerata 117
 Chicken red blood cells (CRBCs)..... 23
 Chromatin 16, 92, 93
 Chromosomal sex determination 16
Cimex lectularius..... 41
 cis-regulatory elements 89, 90
 cis-regulatory modules (CRMs)
 BED form 131, 132
 BEDTools 122
 common function 134
 community efforts..... 117
 computational requirements 121, 122
 custom training sets 123–125
 Drosophila species 124
 enrichment test..... 133, 134
 gene transcription 118
 genomes and annotation files 120
 genomic data sets 133
 holometabolous insects..... 134
 individual insect genomes..... 117
 mature transgenic methods 118
 midgut expression 119
 non-CRMs..... 125
 non-protein-coding transcription 118
 REDfly database 123

cis-regulatory modules (CRMs) (cont.)

SCRMshaw software	118, 121, 122
sorting and merging hits	132, 133
tandem repeats	122, 125, 126
training data	121
Community annotation	75, 76
Completeness assessment	60, 61
Covaris fragmentation system	144
CpG island promoters (CPIs)	101
Crustacea	117
Cryptic species	16
<i>Culex quinquefasciatus</i>	34
Cyberinfrastructure approach	104
Cytometer	24

D

Dengue	213
DNA methylation	
bioinformatic analysis	158
detection	142
in eukaryotes	157
sequencing technologies	158
statistical methods	160
Double digest restriction site associated DNA	
sequencing (ddRADseq)	196, 210
Downstream promoter element (DPE)	100–101
<i>Drosophila melanogaster</i>	24
genomic data	107
promoter architecture	100
RAMPAGE libraries	101–103
TATA and Inr	101
DynaMag-2 magnet	201, 208

E

EggNOG	134
Endopolyploidy	16, 20
Endosymbiont	195
Enhancer	89, 90
Epigenetics	45, 152
Epitect Kit	145
Ethylene diaminetetraacetic acid	188
Euchromatin	16
Expressed sequence tags (ESTs)	100

F

FASTQC plots	143, 148–150, 161, 173
Flow cytometry	
genome size	16, 21
histogram and cytograms	19
internal	19
Fluorescence in situ hybridization (FISH)	
BAC clone DNA probe	192
chromosomal arms	193

chromosome preparations	180
Culicidae family	177
DNA labeling	180
human malaria	177
mitotic chromosome (<i>see</i> Mitotic chromosome)	
neutral-buffered formalin	193
nick-translation DNA labeling	187, 192
polytene chromosome (<i>see</i> Polytene chromosome)	
primer DNA labeling	186, 187
probe preparation	181
protocols	179
repetitive DNA fractions	181, 187
Formaldehyde-assisted isolation of regulatory elements (FAIRE)	
advantages	95
batch effects	95
buffers and reagents	91
chromatin preparation	92, 93
<i>cis</i> -regulatory elements	89
developmental samples	90
DNA binding	90
DNA fragment size	93
<i>Drosophila</i> imaginal discs	91, 92, 95
equipment and consumables	91
experimental methods	89
genomic DNA	96
genomic sites	90
open chromatin	95
preparation, FAIRE DNA	93, 94
sub-optimal crosslinking	96

G

Galbraith's buffer	17, 20
Gene annotations	76, 78
Gene modeling programs	46, 47, 53
General Feature Format (GFF) files	69
Genome annotation	71, 72, 118, 120, 121, 128
Genome assembly	
allpaths LG	4
long read sequencing	2
mammalian	2, 4
primary software	4
strategies	6
validation and chromosome level	7
Genome project strategy	
analysis and publication	6, 7
assembly practices	8
automated annotation	5
bacterial contaminants	4
community workspaces	12
DNA isolation and sequencing	3
genome sequencing	8
human curation	5
insects	12

public database 5, 11
 strategy choice 2, 3
 Genome sequencing 7, 10
 and assembly practices 3, 7, 8
 bacterial 7
 Illumina's NovaSeq 11
 GenomicConsensus package 36
 Genomics 45, 76
 10× genomics 3, 4, 7, 12
 dipterans and hymenopterans 63
 high-throughput nucleotide sequencing 59
 insects 10–11
 resources 67, 72
 GFF3toolkit 76–79, 81–85

H

Heterochromatin 16
 High molecular weight DNA (HMW DNA)
 in agarose plus 30
 DNA isolation 29
 gene finding 27
 genome assembly 27
 insects 30
 nuclear isolation and lysis 28
 nuclei embedding and lysis 30
 nuclei isolation 29
 plug molds 30
 restriction enzymes 31
 re-suspend final pellet 31
 High performance liquid chromatography
 (HPLC) 158
 HiSTAT2-StringTie pipeline 56
 Holometabola 117, 134
 Hybridization solution (Hyb) 219

I

Illumina-compatible library 207
 Illumina primer sequences 200
 Illumina sequencing
 NGS 201
 PCR DNA fragments 196
 SBS 208
 Illumina TruSeq PCRfree Library Prep Kit 146
 IMMBoost method 136
 InParanoid 134
 Insect genomes 3, 10–11
 Insects 75, 76
 Inter-model errors (Emr) 79
 International Nucleotide Sequence Database
 Collaboration (INSDC) 5
 Intra-model errors (Ema) 79
 IPython 51, 55

J

Java 142

K

KAPA Library Quantitation Kit 208
 Klenow enzyme 187
 Kmer content 150

L

LiftOver 136
 Long range restriction mapping 27

M

MAKER pipeline 46
 Malpighian tubules 22
 MasterPure™ DNA purification kit 144
 MCuNovo program
 acquisition 54
 alignment of sequences 49
 BLAST 51–52
 development 48, 49
 gene naming 55
M. sexta and *D. citri* genome assemblies 50
 materials 49–50
 MCuNovo Gene Selector 54
 principles 46
 vs. protein sequence 50–53
 reference sequences 51
 working scheme 53, 55
 Merge phase 84
 Merge pipeline 78, 79
 Methylation 152, 173
 MethylKit 171
 MinHash Alignment Process (MHAP) 36
 Miniasm assembly 35
 Mitotic chromosome
 BAC DNA clones 189, 191
 species of mosquito 184, 186
 3-[*N*-morpholino]-propane sulfonic acid
 (MOPS) 16
 Mosquito larval oral feeding assays 223
 Multilocus sequence typing (MLST) 196, 197,
 199, 200
 Myriapoda 117

N

Nanopore data 35, 42
 Next-generation sequencing 33
 NIH-funded modENCODE project 118
 Non-model insects 118, 133
 Nuclear isolation buffer (NIB) 28

O

Oligonucleotide 198
 Open chromatin 95
 Optical mapping 27
 OrthoDB 134
 Oxford Nanopore Technology (ONT) sequencing 10

P

PacBio data 36, 42
 Pacific Biosciences 11
 Palaeoptera 117
 Paired-end analysis of transcription (PEAT) 101
 Paraneoptera 117
 ParsEval 77
 PBJelly 12, 37
 PCR-enrichment 200, 205
 Photomultiplier tubes (PMT) 18
 Physical genome mapping 177
 Picard tools 143
 Ploidy 18, 20
 Polyneoptera 117
 Polyploidy 16
 Polytene chromosome
 anopheles larvae 182, 183
 half-gravid anopheles females 183–185
 standard FISH protocols 188–190
 Prehybridize 191
 ProLong[®] Gold antifade reagent 181
 Promoter architecture 100
 Propidium iodide (PI) 16, 20
 Proteinase K 30
 Pseudoscorpion 22

Q

QIAquick[®] PCR purification Kit 180, 186
 qRT-PCR machine 216, 218, 223, 224, 229

R

Real-time PCR amplification 209
 REDfly database 121, 123
 Regulatory genomics 118
 RNA Annotation and Mapping of Promoters for Analysis
 of Gene Expression (RAMPAGE)
 filtering and alignment 107–108
 GoRAMPAGE 107–108
 paired-end TSS 108
 promoter identification 108–112
 sequence data, NCBI 105–106
 summary 113
 TSS profiling information 112
 RNA interference (RNAi) 214, 215
 RNase A 30

S

Saccharomyces cerevisiae 215–217, 220
 SAMtools 143
 Sanger sequencing 196
 Scaffolding
 dovetail and 10× assembly 38
 PBJelly 37
 redundant duplication 38
 software 34
 SCRMSHAW software 119,
 121–137
 Seqmonk 162, 171
 Sexual dimorphism 16
 Short hairpin RNA (shRNA) 214, 215, 217,
 219, 220, 222, 226
 Short read archive (SRA) 5
 Single feature errors (Sfe) 79
 Small interfering RNAs (siRNAs) 214, 215

T

Tannins 19
 Target-enriched endosymbiont sequencing (TEEseq)
 adapter-ligated library 200
 alphaproteobacterium 195
 barcodes and Illumina-compatible adapters 204–206
 beads 211
 EcoRI, NotI and ligation of Illumina-compatible
 adapters 200, 203, 210
 enriched library 200, 201, 207, 208
 genetic studies 196
 Illumina-compatible library 205
 Illumina sequencing 201, 208
 iProof DNA polymerase 209
 materials and equipment 198, 199
 microcentrifuge tube 201
 MLST approach 196
 PCR DNA fragments 196
 PCR products 201
 PCR reactions 210
 phosphothiolate bonds 210
 restriction enzymes 209
 Sanger sequencing 196
 schematic overview 197
 T4 DNA Ligase Buffer 210
 validation of library 201, 208
 Wolbachia gene 196, 198–200, 202
 Thermobrite machine 183, 185, 189
 ThermoBrite[™] Slide Denaturation 180, 181
 Thermocycler 203, 205–207
 TopHat-Cufflinks pipeline 46
 Transcriptional gene regulation 118
 Transcription initiation 100
 Transcription start regions (TSRs) 101

Transcription start sites (TSSs)
 demonstration 105
Drosophila melanogaster 108
 genome-scale 99–100
 GoRAMPAGE 113
 hardware 104
 insects 100–101
 materials 103–104
 operating system 104
 RAMPAGE 101–105
 R packages 105
 symlinks 106–107
TSRchitect 104, 105, 108

Transcriptome shotgun assembly (TSA) sequence
 database 63

Transcriptomics 45, 63, 72

Trimmomatic 143

Trizol reagent 218

U

Underreplication 16

Unmethylated cytosines 158, 159, 163, 164

U.S. Department of Agriculture (USDA) 85

V

Vector 213, 214, 217,
 219, 220, 226

VectorBase 63, 71

W

Whole-genome bisulfite sequencing (WGBS)
 bioinformatics 145–153
 biological processes 141
 bisulfite sequencing 142
 canonical CpG dinucleotide context 154
 DMR identification software 155
 DNA methylation 141
 FASTQC plots 149
 hardware requirements 142
 high-throughput sequencing technologies 142
 library preparation 142, 144–146
 PhiX lambda phage 154
 quick installation guides 142–144
 samples 154
 software tools 142–144
 system options 142
 TapeStation gel electrophoresis 146

Wolbachia
 endosymbionts 196
 gene-specific PCR 199
 high-resolution characterization 198
 MLST genes 197
 PCR enrichment 202
 PCR and Sanger DNA sequencing 196

Z

Zika virus 213