Thema: Illumina Sequenzierung

## Name: Dang Quynh Tram Nguyen

### Laden Genome und ChIP-Seq File herunter

(auf NCBI)

* Datensatz DWV\_A Inoculum in SRA Format
* Referenzsequenz Defromed Wing Virus (DWV)

### Konvertieren in FASTQ aus SRA

(nutzen fastq-dump aus sratools)

Code: fastq-dump –split-files <SRA File>

(paired-end reads in 2 getrennte Files)

### Mappen mit bowtie2

* Bauen Index auf

Code: bowtie2-build <Genome-File in fastq> <prefix des Indexes>

* Mappen

Code: bowtie2 -x <Index> -U <fastq File> -S <sam Output>

### Prüfen die Qualität der Daten mit FASTQC

(Graphik zu „Per base sequence content“)

Code: fastqc <sam Input>

„Per base sequence content“ stellt den Anteil der Base in jeder Position des Reads in einer Datei dar. Die Anteile der Basen in Position 1-10 sind nicht stabil, weil es von Random Hexamer Priming bei der Vorbereitung gewirkt.

### Visualisierung zu Positionen von Reads auf dem Genom

* Konvertieren SAM zu BAM File

Code: samtools view -h -S <sam Input> -bo <bam Output>

* Sortieren BAM File

Code: samtools sort <bam Input> -o <bam Output>

* Konvertieren BAM zu BEDGRAPH File mit genomeCoverageBed von bedtools

Code: genomeCoverageBed -ibam <bam sortiertes Files> -bg > <bedgraph Output>

(bedgraph File mit der Anzahl der Reads in jeder Position)

* Visualisierung
  + Auf IGV
  + Auf RStudio