QUALITY CONTROL

Data:

* arm.fastq
* arm.bam
* FL1-1.sam

We will use a tool called **fastqc** to do some preliminary quality control of our raw data. This is a tool that generates summary statistics of sequence and quality data and can be used to filter, reformat and trim next-generation sequence data.

Make a directory in your home directory called fastqcTest.

Getting help:

fastqc -h

To generate report graphs, type the following:

fastqc -verbose –o fastqcTest –noextract –f fastq arm.fastq -

SAMStat is an efficient C program to quickly display statistics of large sequence files from next generation sequencing projects. When applied to SAM/BAM files all statistics are reported for unmapped, poorly and accurately mapped reads separately. This allows for identification of a variety of problems, such as remaining linker and adaptor sequences, causing poor mapping. Apart from this SAMStat can be used to verify individual processing steps in large analysis pipelines.

Use samstat to get statistics on the mapped data file in

smo.bam and FL1-1.sam