

EMBL-EBI

Biological information infrastructure for research

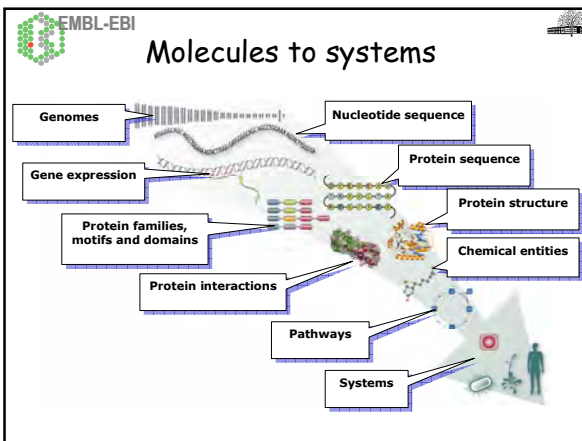
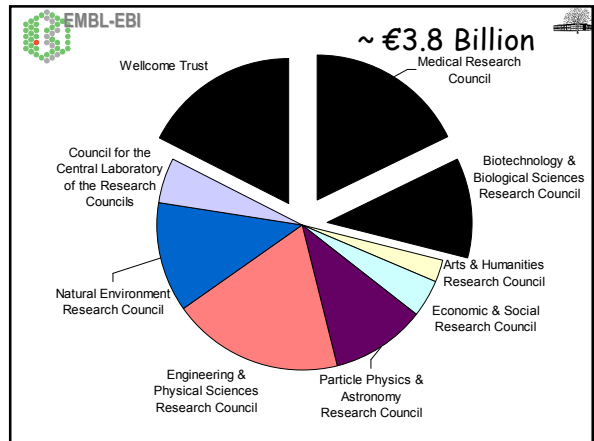
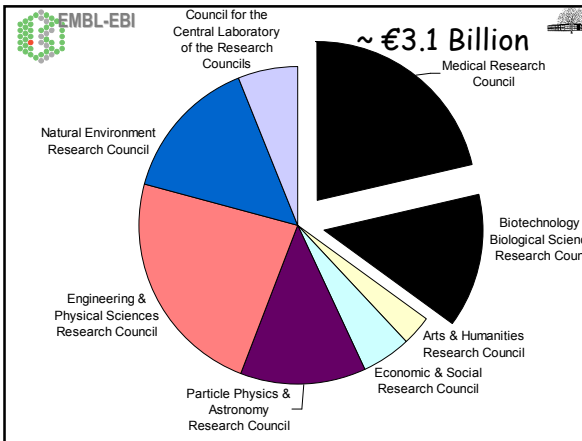
Graham Cameron

EMBL-EBI

From Molecules, to Cell, to Organisms, to Physiology

Genome → Protein (p53 tumoursuppressor) → Cell → Embryo

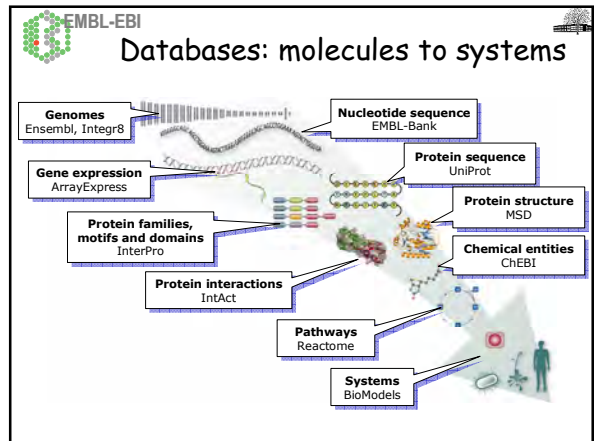
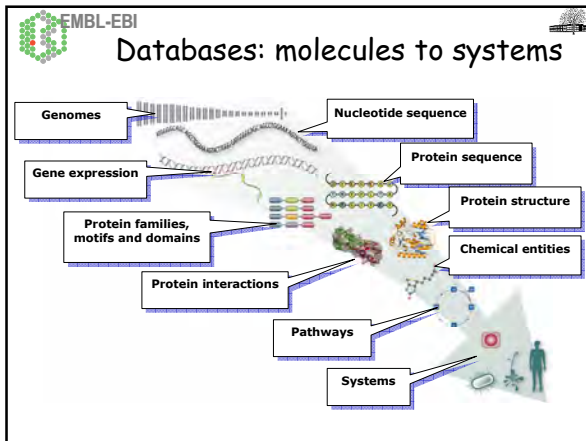
Fruitfly Mouse Human Development, Ageing Disease



EMBL-EBI

We have amassed a wealth of knowledge about the molecular processes of living systems

- Biomacromolecules
- Biologically active molecules
- The behaviour and interactions of these molecules
- The phenotypic effects of molecular changes
 - Mutations
 - Drugs
 - Nutrients
- The molecular adjuncts of phenotypic changes
 - Disease
 - Aging
- Databases
- Web access
- Tools to explore the information
- Systems to capture the information
- Service centres



EMBL-EBI DNA

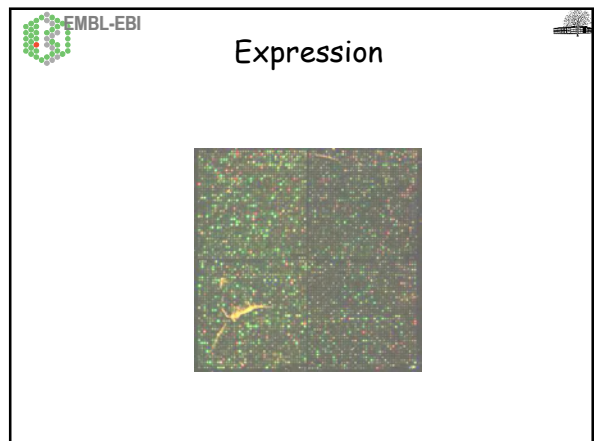
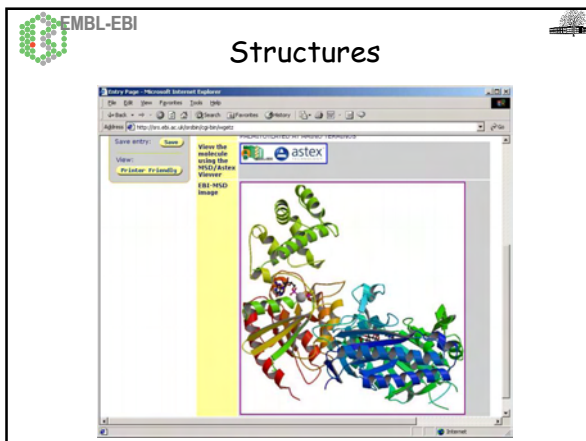
```

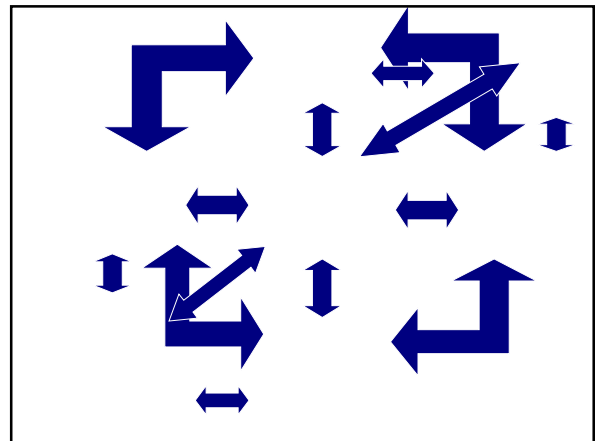
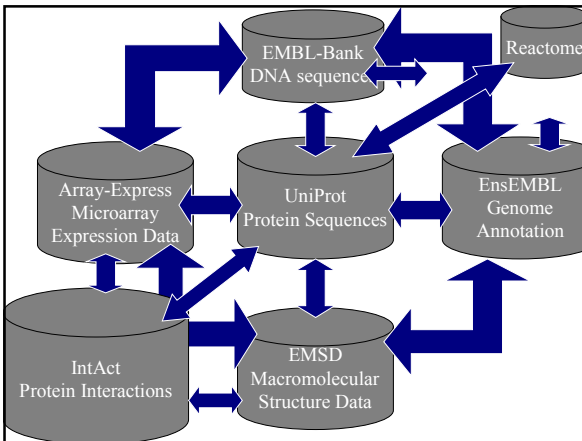
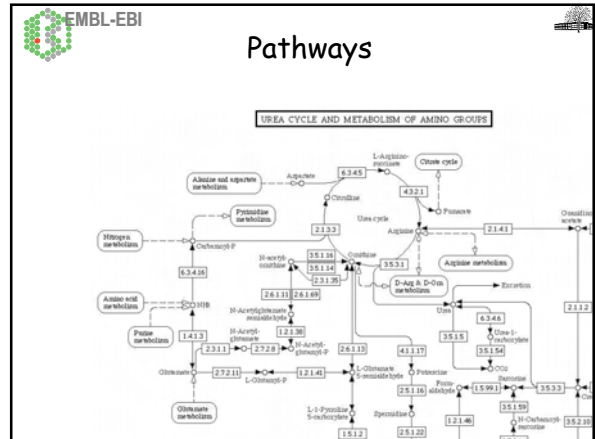
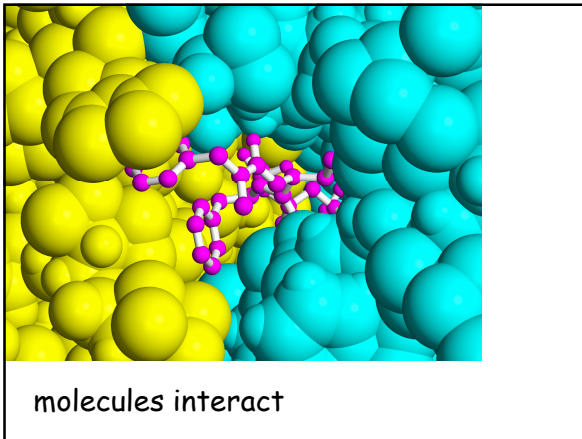
ccctgtgtg  tctctgtgtg  cctctgtgtg  etggcctcc  acgaggcgg  gcccccgc
cgtctgccc  acctgcgct  gctgagggc  gaggctcgg  tcaactcgt  catggcgg
ctctgaacc  gccgcgctt  ccaccaggc  cacatggcc  tgactgcta  cggctcca
cggctggtc  tggcctgca  ggtggtggc  etgctgtgc  ccaagccgg  cagcgcct
gaggcatct  ggtggacct  gttcttcac  tacaccatc  acaagctgt  gccctgac
atggggccc  cctctccag  cggagtgtc  ctgtggccc  tgaacctgc  catcgccc
cggccaaac  ccaagaccg  gttctgtct  aagagctgc  tctccaatg  cctcattt
tcttgacaa  acctgtggg  tctctgacc  cactaacgg  ctgagcttc  ccagagac
gctctcaag  agaccggga  gttctccag  gcccgcttc  actgcaacg  ggagaacc
caacaggac  gctctctct  gctctctct  ccccgcaag  ttgcctgga  gatgaag
gacatcaat  ccaagcagg  ggatatgat  ttcataaga  ttacatcca  gaacatg
aacgtgaca  tctgttttc  tgacatcag  ggtccaaca  gcttgccat  ccagtcca
ccccaggag  tggatcagc  gctcaatgag  cttctgccc  gcttgacaa  gctggctg
gagatcact  gtttaagat  taagatcct  gggatgctt  atactatgt  cctgagpc
cctgaagca  gggccgaca  cgcaccatc  tggatgaga  tggccatga  catgcttg
gcaatctct  tggctggga  gttgcaagg  gtagaagta  acatgcgct  gggatctt
aggggccc  tacactggg  tctctctgt  ctcaggagt  gccagtcca  cgtctggt
  
```

EMBL-EBI Protein Sequences

```

QQQQQHEQQ  HEQQQHEQQ  HVQQQRLAK  RNRGDDPPL  GGDDPLAGP  GFSPRSRAV
QERGGDDCG  GSRRRRGA  GGSSRAPP  OGQGFPAAG  GAEVPRSV  LQDERRRG
RAPEPEAEA  GAPGGDGA  DGDGAPGA  CCRALLOIF  SKRFPSEKL  RLYGRVFFL
NGSLTLRMA  VLVLCVLM  AFHARFPL  LPHLAVLAA  VGVILVAVL  CHAARFROH
MGLACTALIA  VLVAVQVGL  LIPQPSASE  CHMTVPFTY  TITVLLPVM  FAATLSGULL
SALHLAIAR  ANAQDFLLK  QLVSNVIFS  CTNIVGVCH  YPAEVSQRA  FQETRECIQ
RLHSQRENOQ  QERLLLVLP  RRVAMENKA  INAKQEDMF  HKIYIQKDN  VSILFADIEG
FTSLASQCTA  QELVTLNEL  FARFDLAAE  NLCRLKILG  DCYCVSGLP  EARADHAHC
VEMSHDIEA  ISLVEVTV  NHRHVGHS  GRQGVGLG  KRQRDQSN  DVTLANHHA
GGKAGRHIT  KATLSVING  YEVEPGGGE  RNVLRKRSI  EYFLRLCTQ  KRREKAMIA
KMRQRINSI  GHNPPHWAE  RFFYNLGG  QVSKENKRG  FEDPKNAQ  ESANPEVD
EFLGRAIDAR  SIDRLRSEH  RFFLLTFEP  DLKRYKQV  DDFRGATVAC  ASLVFLFIC
VQITIVPRV  FMLSFLTCF  LLLLVFVFS  VIYCVKLPF  GPLOSLSRK  VRKSTNSTLV
GFTTITLVFL  SAFPHEFMC  SEDLGLLAD  ERHSTYSRN  ACHAKASAM  LSLQGGKFC
GTPPSCNFP  EYFTYSVLS  LLACEVFLQI  SCIGKLVLM  AIELIYVLO  EVPRVTLFN
ADLVATANAI  DFNNNQTSQ  CEPATKVAL  KVPTLIISV  FVLALYHAQ  QVESTARLF
LWLQATEEK  EMEELQAVN  RLLLNILPK  DVAAHFLARE  RRNDELYQS  CECEVAHFAS
IANFSEPTYE  LEANNEGEC  LRVNLIIAD  FDEIISDRF  RQLEIKITG  STYMAASGLN
  
```





- EMBL-EBI
- ### Usage
- Basic research
 - Industry
 - Pharma
 - Diagnostics
 - Medical device research
 - Personal care
 - Nutrition
 - Agriculture
 - Forestry
 - Fishery
 - Patent searching and provenance

- EMBL-EBI
- ### European Context
- BioSapiens
 - EMBRACE
 - ENFIN
 - (and many others)

EMBL-EBI

Biosapiens

- European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, Cambridge, UK.
- European Molecular Biology Laboratory, Heidelberg, Germany.
- German National Centre for Environment and Health, Neuherberg, Mönich, Germany
- Université Libre de Bruxelles, Brussels, Belgium
- Consejo Superior de Investigaciones Científicas, Madrid, Spain
- Institut Municipal d'Assistència Sanitària, Barcelona, Spain
- Genome Research Ltd, Hinxton, Cambridge, UK.
- Max-Planck Institute for Informatics, Saarbrücken, Germany
- The Hebrew University of Jerusalem, Girat Ram, Israel
- Department of Biochemical Sciences University of Rome "La Sapienza", Rome, Italy
- University of Stockholm, Stockholm, Sweden
- University of Oxford, Oxford, UK.
- University College London, London, UK.
- Radboud University Nijmegen, Nijmegen, The Netherlands
- Swiss Institute of Bioinformatics, Geneva, Switzerland
- Technical University of Denmark, Lyngby, Denmark
- University of Helsinki, Helsinki, Finland
- University of Geneva, Geneva, Switzerland
- Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary
- University of Cologne, Cologne, Germany
- Institut Pasteur, Paris, France
- BioInfo Bank Institute, Poznan, Poland
- Max Planck Institute for Molecular Genetics, Berlin, Germany
- Genoscope, Evry, France
- University of Bologna, Bologna, Italy
- European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, Cambridge, UK

EMBL-EBI

EMBRACE

- European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, Cambridge, UK.
- European Molecular Biology Laboratory, Heidelberg, Germany.
- Institute of Biomedical Technologies, Section Bari, CNR, Bari, Italy
- University of Manchester, UK
- Swiss Institute of Bioinformatics, Geneva, Switzerland
- Swedish University of Agricultural Sciences. The Linnaeus Centre for Bioinformatics, Sweden
- Centre National de la Recherche Scientifique, Clermont-Ferrand and Lyon, France
- Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark
- Centro Nacional de Biotecnología/Consejo Superior de Investigaciones Científicas, Madrid, Spain
- University of Stockholm, Stockholm Bioinformatics Centre, Sweden
- Institut National de la Recherche Agronomique Toulouse, France
- Max Planck Institute for Molecular Genetics, Berlin, Germany
- CSC, the Finnish IT Center for Science, Espoo, Finland
- University College London, London, UK.
- The Weizmann Institute, Rehovot, Israel
- Centre National de la Recherche Scientifique, Clermont-Ferrand and Lyon, France
- Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark
- Carretera de Ajalvir, km. 4, 28850 Torrejon de Ardoz, Madrid

EMBL-EBI

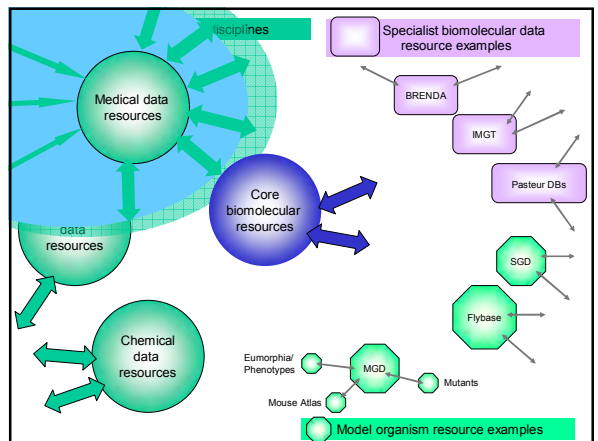
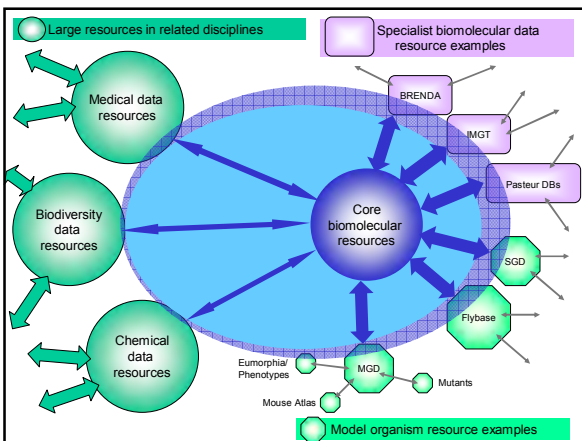
ENFIN

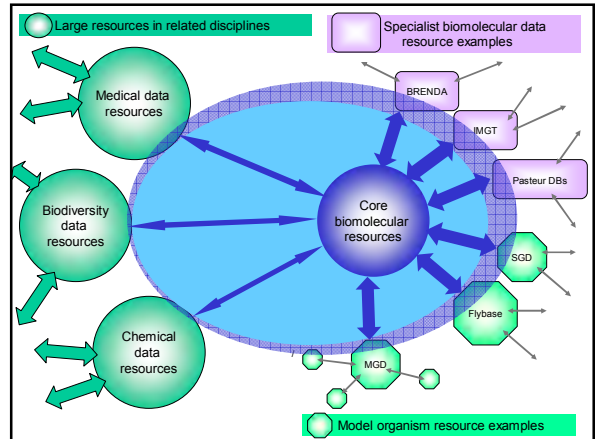
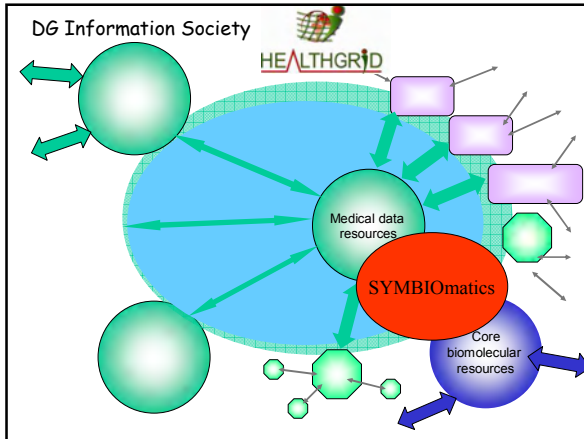
- The European Bioinformatics Institute / The European Molecular Biology Laboratory, Europe
- The University of Dundee UK
- Technical University of Denmark
- University of Rome Tor Vergata Italy)
- Medical Research Council Mammalian Genetics Unit (MRCMGU), UK
- Ludwig Institute for Cancer Research, Uppsala (LICR-UPP), Germany
- The Max Planck Institute, Germany
- University of Helsinki (UH), Iceland
- University College London (UCL), UK
- National Center for Research and Technology, Hellas (CERTH), Greece
- Universitaet zu Koeln (UNIK), Germany
- Weizmann Institute (Weizmann), Israel
- Egeen (EGEEN), Estonia
- Serono Pharmaceutical Research Institute (SPRI), Switzerland
- Consejo Superior de Investigaciones Científicas (CSIC), Spain
- Centre for Integrative Bioinformatics VU (IBIVU), Netherlands

EMBL-EBI

Global Picture

- DNA – tripartite international collaboration (including patent data acquisition)
- Protein sequences – Uniprot collaboration
- Macromolecular structures – tripartite international collaboration
- Intact international agreements
- Reactome – USA Europe collaboration
- Etc.





EMBL-EBI

Using the information

| | |
|-------------------|-------------------|
| Healthy | Diseased |
| High Yield | Low Yield |
| Disease Resistant | Disease prone |
| Salt Tolerant | Not Salt Tolerant |

Suppose a gene's variation seems important

EMBL-EBI

Using the information

| | |
|-------------------|-------------------|
| Healthy | Diseased |
| High Yield | Low Yield |
| Disease Resistant | Disease prone |
| Salt Tolerant | Not Salt Tolerant |

Look in databases for similar genes, their products, and functions, structures, interactions and expression patterns. The processes in which they are involved.

EMBL-EBI

Using the information

| | |
|-------------------|-------------------|
| Healthy | Diseased |
| High Yield | Low Yield |
| Disease Resistant | Disease prone |
| Salt Tolerant | Not Salt Tolerant |

Can we influence the processes in which they are involved?

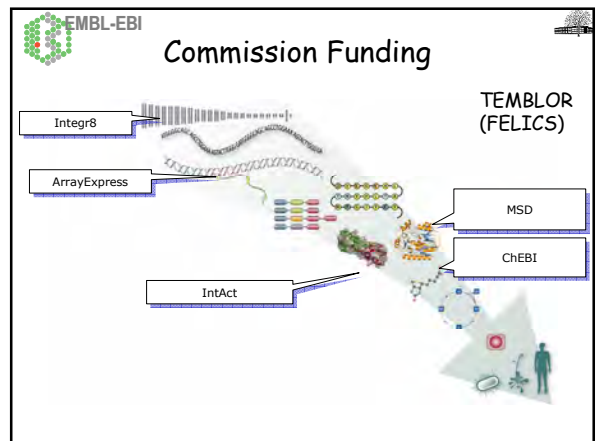
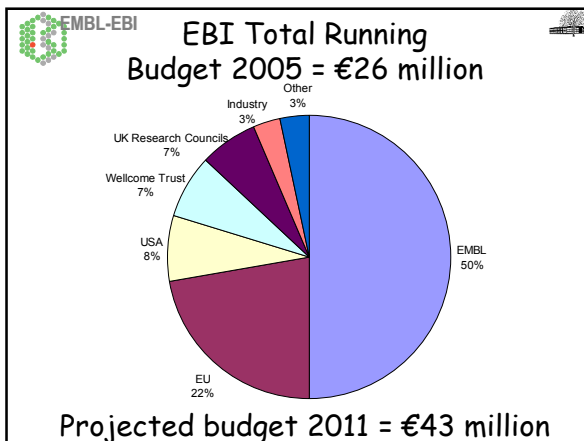
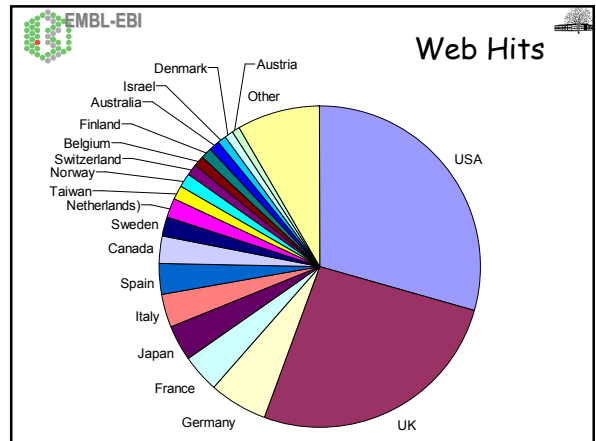
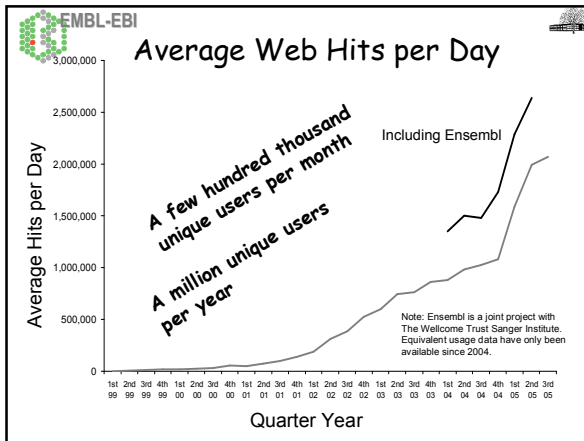
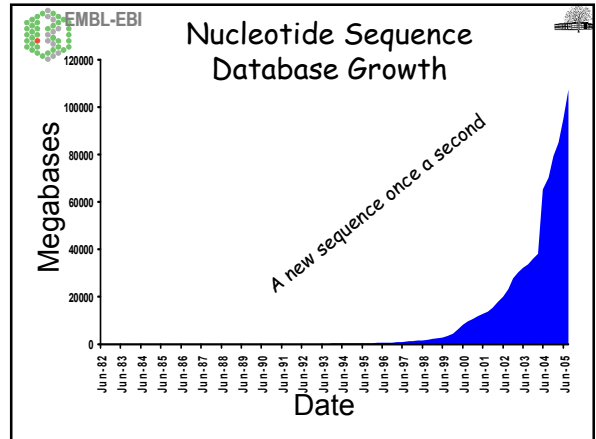
EMBL-EBI

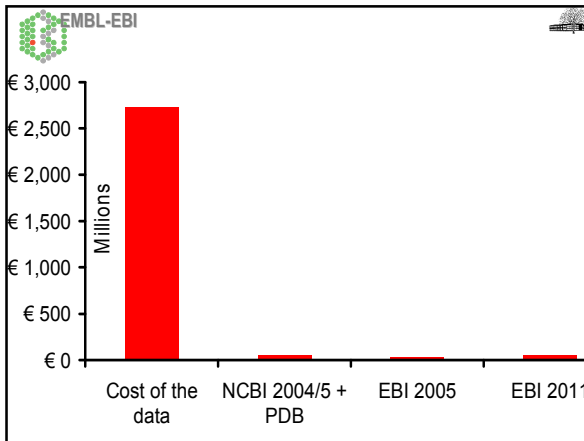
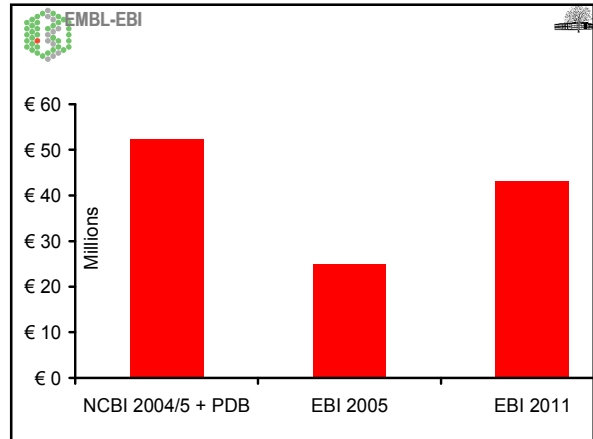
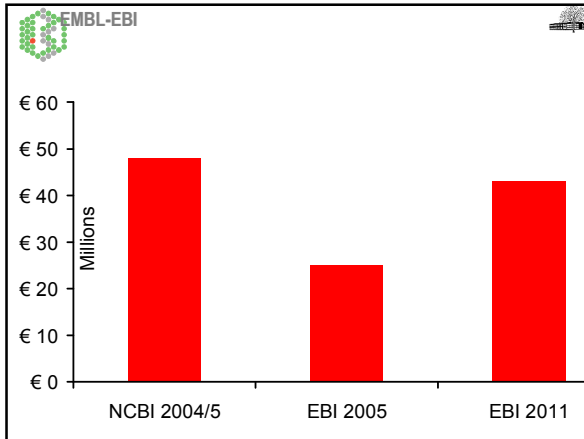
Using the information

| | |
|-------------------|-------------------|
| Healthy | Diseased |
| High Yield | Low Yield |
| Disease Resistant | Disease prone |
| Salt Tolerant | Not Salt Tolerant |

Can we influence the processes in which they are involved?

- EMBL-EBI
- Working out what in the lab what a gene does could easily be a year's work
 - Searching databases can do it in half an hour





EMBL-EBI

Today's most optimistic costs

| DNA Data | |
|-----------------------------|-----------------|
| Cost per read | € 1 |
| Read length | 500 |
| Coverage (X) | 5 |
| Cost per megabase | € 10,000 |
| Megabases (EMBL) | 107,563 |
| Database cost | € 1,075,625,807 |
| Structure Data | |
| Cost per structure | € 50,000 |
| Structures in PDB | 30000 |
| Cost of PDB | € 1,500,000,000 |
| Microarray Data | |
| Cost per hyb | € 500 |
| Hyb's in ArrayExpress | 30000 |
| Cost of ArrayExpress | € 15,000,000 |
| Percentage of data captured | 10% |
| Value of all data | € 150,000,000 |

EMBL-EBI

Bioinformatics Infrastructure

- Has captured the data from several billion Euros worth of science
- Serves a community of perhaps a million users
- Supports science on which the UK alone spends €3-4 billion a year
- Cuts years of lab work down to hours of computer work
- Is crucial to human well being from medicine to agriculture
- Sees data volume and usage growing exponentially
- Might cost a few tens of millions (at most a couple of percent of the cost of the science it supports).