DEPARTMENT OF HEALTH & HUMAN SERVICES

National Institutes of Health
National Human Genome Research Institute
FOIA/PA Office, RKL 1, Suite 6070
6705 Rockledge Dr, MSC 7957
Bethesda MD 20892-7957

April 22, 2011

Robert Cook-Deegan, M.D.
Duke University
Institute for Genome Sciences
& Policy
Durham, NC 27708

Re: FOIA Case Number: 11-FOI-00165-NHGRI - 38639

Dear Dr. Cook-Deegan:

This is our final response to your March 3, 2011, Freedom of Information Act (FOIA) request addressed to the National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH). You requested: 1) any notes (hand written or otherwise) taken at the 1996 International Strategy Meeting for Human Genome sequencing, and 2) any email or memo summaries of this meeting distributed within National Institutes of Health afterwards.

We searched the files of the NHGRI Office of the Director for records responsive to your request. That search produced 48 pages responsive to your request. Enclosed are: 1) notes taken at the 1996 International Strategy Meeting for Human Genome sequencing (45 pages), and 2) email/memo summaries of this meeting distributed within National Institutes of Health following the meeting (3 pages). A total of 48 pages are being released with this response.

In certain circumstances provisions of the FOIA and Department of Health and Human Services FOIA Regulations allow us to recover part of the cost of responding to your request. Because the cost is below the $25 minimum, there is no charge for the enclosed materials.

Thank you for your interest in the National Human Genome Research Institute.

Sincerely,

Christy Cecil

Christy Cecil
Freedom of Information Specialist, NHGRI

Enclosures – 48 pages

Wellcome Sequencing mtg. - Bermuda 2/25/94

I. A Morgan -
Cooperation, created action
B Watson -
10th anniv of strategy discussions Los Alamos 3/86
Cooperation in technology & cost, not duplicating efforts
C. Presentations
1. Waterston
Regional STSs or large insert BAC lib (10x)
↓
Local map construction to select
minimal tiling path
↓
Seq. & mixed shotgun/directed
↓
Annotate, submit
Don't miss the chance for a high quality sequence
& continuity & high accuracy (99.99%)
Goals 50Mb in next year for Waterston/Wilson team
Aim at chr. 22 completion
Also X, 7, 6
May do some targeted effort too

2. Lander
Anchored BACs (& STSs)
Shotgun & seq-directed closure (automated)
Distinctive focus:
Front end automation (STS mapping of BACs, characterization &
select | Separation
Hands off assembly

5Mb in year 1

3. Venter

Chr 16 STSs → BACs → cg

BAC end sequencing:

5x insert ≥ 150Kb (99.9% coverage) → 300,000

End sequence → 600,000 sequences (10%)

Refine Hot seed clone <1 yr. ~ 32 ABI's

→ Find next one in Hbag path

4. Hood

Olson MCD

B. Green Assembly

BAC's 4 restriction digests No new reagents

Have 100% success in first-sequencing 35 BACs

500 bp reads

Repetitive seqs? Alus are no problem

LINEs are a problem

11% of ends were this

Strategy also gets ~ete cDNAs

Allows multiple participants

5. Hattori – subst. Ogasawara

Japanese effort –

Monbusho will give new 5-yr. effort "Human Science"

2 new labs

STA – Aim for 5Mb/yr. Nakamura, Sakaki, Shimizu, Ikeo

6. Cosrano

Amount 19 cosmid map + EcoRI maps
50 Mb euchromatin — have 48 Mb
Started sequencing a 20 Mb region
Shotgun & directed closure
Lots of Alus
LLNL recognized & sequence as the donor
Focus on repair genes

7. Gibbs - X 12

6 institutes, 25 people
Reduce redundancy & increase sequencing
Focus for 5 Mb Xq28, Xq22 Aim for 2c4/bp

8. Moyzis

85% of chr. 16 in cosmid / BACs 100 Kb STS map
chr. 5 - 100 Kb STS (Mich + MIT)
DOE funded at 1X SANGER
Move on & only 70% sequence
98% chr coverage
Have have 2 Mb of 16p
Lots of EST hits

9. Lehrach - German genome program

Will be oriented toward functional info
Big influence in research resources center
70% map of X
Working on 21

10. Ansorge
2-dyes, 2 lasers
Internal or primer label
200 clone detector-
Avg read 600 bp - expect 1-1.5 Mb
Expect someone to hand them the clones
Plan to produce machines

11. Weber - Whole genome sequencing
Use long (≥5 Kb) reads and a shotgun
Quality value to each nucleotide
Aim for 10x coverage (30 Gb of raw seq.) by 2000
Pool multiple sources of DNA → get polymorphisms
Myers has done computer simulations on assembly
$0.005/bp raw sequence
×10 = 5¢/bp (Assembly)

I. Mon Ann Sequence - Rudy Myers
A. Sulston → multiclone - need to revisit template to get to only one
C. elegans 7x coverage (~ gaps) 17,500 More in gene-rich regions
YACs complete = 7 gaps
Working on clones of gaps on X then gaps
finished 33.9 Mb 6000 gaps (1/5 Kb abound, 1/7 Kb ×)
45% of gaps show data base match
50% of seq. is intron + exon, 28% coding
Total gene count est. 14,000
1/3 of predicted genes have finding ESTs
Central lets others refrain from using the data before anyone else does

Fingerprinting. Quality is worthwhile
* Error rates in cosmids & M13 are low - acceptable assembly?

Sulston says we should change here a need for
human finishers
Gaps - ① long range PCR
② Other bacterial clones (fosmids, diff. hosts)
shotgun YACs into sequencing (M13, plasmid?)
Completion will be asymptotic

B. Craig Venter
ESTs: 355,610 in TIGR web site
Level 2 17,015 (4.7%)
THCs 51,099
Singleton ESTs 129,769
TIGR Assembler
H. flu
25,000 seq. runs 85-90% completion reads
Predicted 82 gaps
Plasmid templates, do forward & reverse, constrain
assembly by distance
Shotgun is diluted in repet. areas
λ clones to cover gaps 1727 genes
End sequence pairs λ - walk across
22% of coding regions had no detectable matches
Mycoplasma genitalium
470 genes
Web site shows homology that serve to identify gene
42% unassigned
Methanococcus - will publish soon
282 contigs → 14 groups using F/R sequence info
(200 sequence gaps)
↓
unbanded PCR
1900 cons

Data release –

Publish when annotated

Thinks ftp site is a good idea

Strep. pneumoniae — having trouble w/ libraries

Treponema pallidum (NIAID)

DOE –

Wants to be part of Bioinformatics

Mouse/rat WWW & fold/span ?

Estimates 30 ¢/bp (direct) total ~ 50M

Existing resources, how are they being used?

C. Myers – Transposon-based sequencing of minimal tiling path BACs

Chr. 4 20 Mb

Rhesus – C3 – 7300 STSs. → not on www yet. Will be assembled into maps by CSH

TNG Now available → Now 60?. Likelihood of hitting with 1000 more

STSs – 4000 now

3000 requested from others

8–10K/yr. to get 30K (200/wk)

1/2 cDNAs

TNG's tested on a few regions

~~How~~ WWW finished 63 at 10K

Then → TNG (will do some 10K, plus add 20K more)

Southern "There ain't a single published validation of chips for sequencing or sequence validation"

D. Bentley X, 22, 6

5 Mb Xq22
2 mb Xq25
3 mb Xq26
} sequence - ready

Whole chromosome
Flow sorted → 96 - 97% purity (from hybrid)
Fingerprint
Contigs

22, X, 6, 7, 5, 20

Framework map
↓
PACs
↓
Fingerprint
↓
Contigs

merge

Chr. 22 (Wash U/Sanger doing distal to NF2)

Linked at 4x → 10x

60% from whole chr. fingerprint → ~ to 250 Kb

8 Mb ready for sequencing

Chr. 6

Not much by YAC map

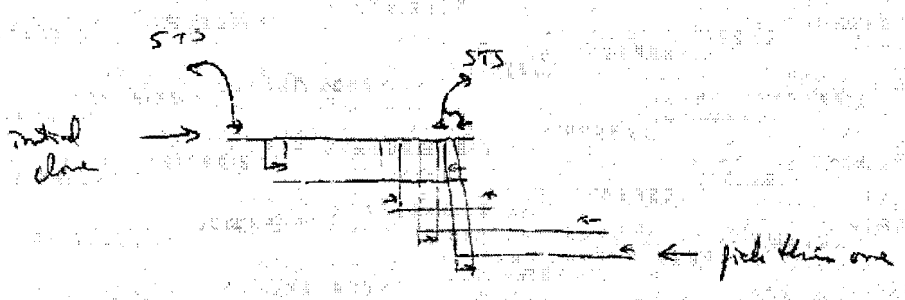
1200 markers in public domain

700 worked 679 RH mapped

Tried F1, on PACs 10 - 20 Mb
(completed)

6 with STS

26 linked to another STS



E. Lander — Resources

Human: 16,500 STS now

8000 RH additional at Whitehead

11500 YAC additional

May go to ~20,000 ends 6/1/96

Mouse: 14× YAC lib
6× BAC lib (F?) — 10× on the way to RG

Rapid STS intact mapping

Hypothesis: to for validation, not for initial tiling path

F. de Jong

PACs — pCYPAC2 ~18

pPAC4 (19.5 kb) EBV, hygR, SV40 prom/polyA

BACs — pBACe3.4 (11.4 kb) — Tn7 to add dom. selectable markers later?

PACs — 16× ~5% empty wells

(pCYPAC2) RsaI 1, 3, 4, 5 digests on 3×, 3×, 4×, 6×

46, ×y

Need moved to BACs — expect 10× by CSH

Will make rec. mouse April/May start

Good so for PACs — might want BACs instead

Stability?

FAB BAC is said to be 120 kb at Genome Systems

220 kb in their hands

Seq'd 350 ends & make Telenius primer PCR

→ gel purification of product

Use cre-lox to introduce Probe labelled ends

G. Sonon

RG BACs >15x by 4/96 — 3 sources (2 are sperm) ⇒ 6 haplotypes

Screen by hybr. 1 probe

Ch. 22 BAC contigs 4x Gb 350Kb avg.

Links EST → BAC maps

Cell lines will be * available (One person will be 15x)

** Informed consent?

M. Evans

17,952 reads - 5.5x coverage

14,727 T7 end seq. → BLASTN/X, 9000 matches → 950 ε P<10^-15

29/174 known genes in chr. 11

8858 hybrid using YAC probe (221)

12,138 clones fingerprinted

10 contigs + 1 PAC seq'd at high accuracy

WT2 region sequenced

Likely candidate

Plan for 2 mb by June

Did 2 contigs entirely by primer walking

"contig-oriented walking" - using high density end-seq.

Gene build 192 clones oligo seq.

PHRAP → direct clone

Cost 5¢/base $1/oligo

LTI

III.

Interlude - Bob & John

✓ ① Timing of Data Release - Immediate (1 Kb contigs)
② No patenting of genomic sequence ∃ functional info
③ Genes should be freely available for research work (Lashkari)

} I think we agreed

Funding agencies are asked to set policy

* Chris Fields asks for Letter to NCGR from NIH - reminding about data release policy
A longer gene hunt talk

A. Hood ~ Phil Green's
PhDQ (Phred + quality name included) → 1/5Kb
ABI is ~10× ↑
Errors are from Phrap, inadequate data, editing
Highest quality areas: 1/167Kb

B. Wilson
Base-calling
GETLANES (replaces ABS) - will be able to do 48 now, later on ~377 - runs on Unix, available now
PHRED+Q
TPP - color separation, baseline subtraction, mobility correction
ABI - upgrades available 6/96?
As of 4/1 - will be at 39Mb/yr finished $0.40/bp

Q: Does that reverse reads aren't that helpful
(but this is for contigs)

C. Sulston (for Rogers)

2 mb of HD (1.8 mb actually) - "sequence ready"

Cost ~20p (direct)

PACE - tracking system

Chr-22

Chr-6 - Sanger Centre (Wellcome) will start this summer

Arbo. phage X, 20.

Human assembly 4.2

finished 0.2

submitted 2.7

7.6 public

× 0.40 = 3.6 m

By 97/98 expect ~90 mb/yr. human
(that's current funding)

D. Gibbs

BODIPY dyes - structural similarity → epistatic

2 dyes/primer energy xfer

More chemistry 16× less template

Terminators have been constructed

Annealing reagents needed & Mathies primer chemistry

* How will they be made available?

E. Hawkins

40,000 templates/d to do 3 Gs in 10000d

Solid phase reversible immobilization

Separation - DNA purif, seq set up, thermal cycling

6 hrs/20 plates/1 FTE 8000 samples/24 hrs

No user intervention

Flexible - M13, PR, plasmid

Could go to 12,000

CRS attended ann ... commercially available. Without reality
4mm track
120 plate carousel
Running x 6 mos
15 - 20 plates/day
M13 subclones (2kb) → 4K Dye Primer
pUC (5kb) — 1.2x K/R Dye Terminator

Finishing strategy — FINISHER, hands off automated
Bar code encryption
LabBase data management

GRACE — 6K analyzer. Monitors Phred good scores 0-3
5mm to 4K linear gel
Running 50 lanes

F. Rosenthal — Jane
36 people (15 techs, 7 PhD, 12 postdocs, 2 informaticists)
12 ABI 377
1 Licor

Strategy — Shotgun Goal → BACs
100% dye termination - better GC rich
Closure - long + some walks
Directed primer walks

PCR amplification of purified clones from cDNA pool (5-12 kbases)
3 gels/machine/d
Aim for 10-15 mb finished Seq/yr.
Goal: 3 mb in Xq28 Main Fanconi coord only (10%)
3 mb in Xp11.23-24
1-2 mb in PAR1 of X
Second 1-2 mb pieces in 7, 11, 17 (D2 genes)

Ch. 21 Seq. [illegible] (10 groups E to do 30-35 Mb
in 3 years
Also look at mouse & Fugu
Sanger Academic
Sanger Cons.] 250 – 1.5 mb/group/yr.
even on left from yeast
Not targeting the Y mb Down region (II)
Region I
III 5 Mb
IV 0.8 Mb PME region
Fugu - Man base 200 kb
LI - same # exons & size
synteny to chr. 1
G6PD consid - has a dump 1 5 → on chr. 12?!

Cost 1 Mark 60 year 1 (~ 91)
Why so many centers - politics need to have
allies
Lots of ABIs are there already
Expect to drop out the non-performers

G. Roe - chr. 22 den → NF2
Strategy - ds vectors FS, terminators
nebulize for shotgun
Clones - long 12 hr reads
Caltech - molecular out of BACs, 2 BACs done
Putting their data on Sanger
D. George - mouse/human cosmid match up
Fug X region 6 months

remaining deleted region β-adaptin
Ended 2 BACs
Status ~3 Mb done (0.25 Mb in Gen Bank) Will be entire in next 2 mos
One clone was 80% Alu
Recently funded by NIAID to do S. pyogenes ε N. gonorrhea
↓
Load to gel
DNA
↓
10,000 complete isolates
2-3K sequenced

H. Weber
Genotype output 1.2 M/year in 1996
Cost ~$1/genotype by end of 1996 (maybe?)
Total genome shotgun
4.8 × 10⁹/day in 4 yrs
0.55 ¢/bp

* Next step lower Tag price?
Work both PCR & ribosomes

I. Hattori:
Ch. 21 - STA
Mental deletion
1500 PI's (2x)
→ Depot, subclone → mental deletion → Seq. 600bp
Seq both ends run pl to assess use
make map software prods
~400 bp/read

Did 300 Kb of APP 3.4 × redundancy
Now trying to do 1.8 Mb of Down syndrome
→ 14 PI's Finish by end of May

annotation / polymorphism?

Has 4 genes
Plans to do rest of chr. 21 (!?)
Agrees it necessary to do ds
99.7% with SS - they think good enough

IV. Informatics
A. Adams - Annotation

GRAIL
Homology
ESTs
] nice to integrate all of them

Who updates annotation? -
GenBank is willing to
GSDB emphasizes community annotation

B. Hillier
Data must represent sized gaps
Represent polymorphism
Assessing error rates
1) local analysis = annotation locally is helpful
2) Compare to existing entries
3) Alternative assembly ± diff. scheme
4) Confirmation using - STS, restriction, fingerprints, overlapping clones
5) Public availability

Assign a GDB # to every clone?

C. Wilson - Lay clones + small clones
Simulations of whole genome shotgun assess good behavior
EL: Diploid, 10^5 reads to compare to ref sequence should yield 4×10^7 bp $\rightarrow 4 \times 10^4$ polymorphisms

EL: Need a finishing plan that will cut up the factor of 2

D. Durbin

Coding for accuracy –
Phrap/Phred can produce patches
1/31 bp
↓
1/167 Kb
What about the preliminary data? Should also include
Need to retain original bcltfile to have best idea of confidence?
Storing traces? – what to archive?
Recently upgraded to core bcltfile

V. Final Session

A. Data Release /IPR

Release – automatic release of assemblies >1 Kb, probably daily
mandate submission of finished annotated sequence

IPR – Aim to have all human genomic sequence freely available for both research and development, and in the public domain for large-scale sequencing centers
in order to maximize its benefit to society

Enclosed unanimously

Primary means is the absence of established experimental information about function or diagnostic utility

Endorse the principle that genomic sequences, ~~factual~~ in the absence of experimentally any additional information about functional or biological ~~value~~ substance, is not an appropriate subject for patent protection

B. Coordination

Hood (Seattle) – 25 mb of chr. 7 TCRα/β MHC/HLA

Simon / Adams – 30 mb of 16p / 3 yrs

Rosenthal – 30 mb of 21q (excluding DS, PMS) / 3 yrs
+ Xq28 (coord. ε Gibbs)
+ Xp11.23-24 + PAR1
+ 6 more regions of 1-2 mb
Hypophosphatemic rickets

Moyzis – 16 and Sp
↳ 16p part not being proposed by TIGR/Sanger ~20 mb

SASE plans ? amount
Stated plan was 100 mb (all of 16)
20 mb of finished

MIT – Quite vague
105 mb / 3 yrs. Most chr. 17
Heredity from mouse systems

Gibbs – Xq28, Xq22, 12p13
30 mb / 3 yrs.

Roe – chr. 22 NF2 → cen
~~20 mb~~ 6 mb

Japan – won't say

Chen – 4 mb of X

25 mb

Carrano – chr. 19 50 mb
Rate unclear 4 - 10 mb / yr.
chr. 2 later

St. Louis – 90-100 mb / 3 yrs. (wanted 350)
chr. 7, 22, X

Melanie - chr. 13/18 5 mb (of funded)
Gait seq throughput

Rogers 550 mb/7yrs. 150-~~250~~ mb/3 yrs.
22, X, 6, 20, 1, 16p

~~600~~ —

Weber A few mb of whole genome

Evans 11p15.5, 11p12, 11q23
2 mb in each → 6 mb/2yrs
+ sampling

Kucherlapati
?25 mb of chr. 12

LBL - Chr. 5q 12 mb/3 yrs.

Stanford - chr. 4

C. Funding Agencies

1. NCHGR
2. Wellcome - Space is enough to do all of it
End of June - occupy labs
Cap. fabrication end of 96
Produced 6/97

Pathogen sequencing
Sanger - TB
Malarial/chromosome

GENSFUN (Genes to structure & function)
Wtg. in 9/96
Is UK well poised?
Bioinformatics, Structural Biology, cell biology
protein expression, transgenics

Budget Sanger £91M/7yrs. for all of it → Do 1/3
I £-9 m/... for sequencing $15M/... in 3 yrs.

3. EC

5 yr framework progr —

Now in 4th 1994-98 (FP4)

20 programs $17B US dollars

15 member states

+ now Israel - Norway - Iceland

Biomed - Human Genome Research

$11M/yr.

5 subareas -

mapping/seq -

function

disease

gene Tx

data management

Currently spending ~$3M on seq-related

Advisory group meets Thursday to decide

on 1998-2002 (FP5)

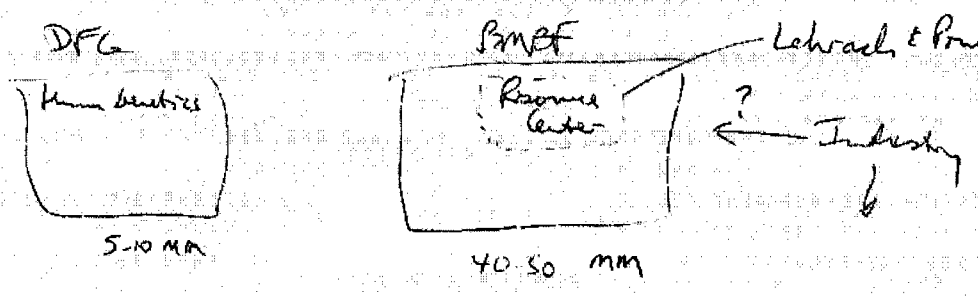
4. French

"Serious discussion" of a large scale sequencing effort

5. Germany - Laplace

Next week Sci Adv. Committee will meet

Some players on here



6. DOE (South)
$70 M/yr
About $10 M is going toward human
Expect to see $40 M/yr in the next 3-4 yrs.

7. MRC
Funds cleaning - got an extra
$ 10M to hang it along

D. Plans
Another outing of same group in 1 year
HUGO has a web site on GDB -
List current plans

E. Notification of clones to be sequenced
Put on a Web site
All of 22 will be on Sanger
HUGO Web page will point to where info is

F. Accuracy
What rules?
ds coverage?
rule of 3? ⇄ or two diff. determinations

Plans - do them at same accuracy? FC said yes. Incremental cost not much ↑
No real consensus - defer to next year?

Gaps - How many?
GC rich regions? } "Technical" gaps
$(CA)_n$?
$(N)_n$
"Biological" gap - not represented in Subclone
PCR might work then
But if it doesn't

Need more research on GC sequencing?

Goal is no gaps
If all the fails, must size, map

Quality check?
Resequencing would require > 100,000 bp
GC content will affect
Try re-assembly?
Thus: ① about resequencing traces
Quality of traces:
Reassembly - problem for non-shotgun?

Should restriction digest be required? 25 ng
Europe opens yes
Not a big extra cost
{ Hind III
Bam
RI

Princess Hotels

NCHGR — Funding climate — Shortfall — New $8.5M → current extramural budget is $128M (See back)

RFA – Tech. Development $6M

Large Scale Pilots – process for review

Awards by 4/1 (we hope)

Review 12/97 (SS recommended)

Council 1/98

Competing renewal 9/98

Real $ $70-80M by then? $20-30M/yr ↓ 10%

Other issues –

Database submission

Timing — Pre-release ? what does it mean
Formal submission will be
GenBank vs. GSDB evaluated
on that

Data representation Accession #, updating

~~Curation~~ : Represent uncertainty identifiers: ? workshop in April
what info to include sources, users,
some of info

Quality checking — is it not irresponsible to ignore?
High quality at the beginning is essential
Does it include comparison of clones to genome?

Tracing review?

Independent resequence?

How to resolve differences?

Yeast RFA

Preparing for after the genome

1) Technology – seq., genotyping
2) Whole genome function

3) Evolutionary Bio –
4) Human variability
5) GLSI

96

#16m Tech Dec. +5 → 21

#14-5m Production +15 → 29-30

NTMC degrees

26m fly

Cdr

RESOURCES AVAILABLE

Name of participant : Mark Adams / Craig Venter - TIGR

Nature of resources available (software, maps, clones etc.)

Software
TIGR Assembler - sequence assembly

HBQCM - hexamer-based composition tool

yank - GenBank extraction software

TIGR's sybase schema

Human cDNA Database - >355,000 ESTs, >51,000 THC assemblies

Available via:

e-mail request to arkerlav@tigr.org (Tony Kerlavage)

cDNA clones through TIGR/ATCC and www

Any conditions attached:

None.

RESOURCES AVAILABLE

Name of participant : ANSORGE/EMBC

Nature of resources available (software, maps, clones etc.)

- SEQUENCING TECHNOLOGY - 100kb/per run
GENESKIPPER - ASSEMBLY PROGRAM
+ SEQUENCE ANALYSIS

RAN-DI (Random-Direct) strategy -
. assembling first 80-100 clones
+ all ECOR1 fragments randomly sequenced
→ NO CLONING GAPS OBSERVED. finish route DIRECT strategy

Available via:

FAX or by mail to EMBC

Any conditions attached:

RESOURCES AVAILABLE

Name of participant :

Tony Carrano
Lawrence Livermore National Laboratory

Nature of resources available (software, maps, clones etc.)

Resource	Availability
High-resolution, metric map of chromosome 19	Published version available in Dec issue of Nature Genetics. Detailed version available by collaboration
Arrayed cosmid libraries of human chromosomes	Through major genome centers. Soon to be available through the UK and German resource centers.
IMAGE collection of cDNAs	Available through industry and resource centers.
DNA sequence sample tracking software	Contact Tom Slezak @ LLNL
Clone fingerprinting assembly and database software	Contact Tom Slezak @ LLNL
Mapping infrastructure resource (creating high-resolution sequence ready maps in cosmids and BACs)	Contact Tony Carrano @ LLNL

Available via:

see above

Any conditions attached:

Creating maps as part of the mapping infrastructure resource would require funding.

RESOURCES AVAILABLE

Name of participant : RICHARD DURBIN

Nature of resources available (software, maps, clones etc.)

SOFTWARE: ACEDB database system

SAM (Carl Soderlund) marker assembly/edit/viewer

FPC [redacted] (Ed a) fingerprint " " " SOON

AUTOEDIT – sequence automatic editor for assemblies (Richard Mott)

MSPCRUNCH/BELVU/DOTTER – sequence analysis/viewing tools

Available via: ANONYMOUS FTP (FTP.SANGER.AC.UK)

email: RD@SANGER.AC.UK

Any conditions attached:

NO COMMERCIALISATION (use by companies OK)

Resources Available

Name of participant:

Glen A. Evans

Nature of Resources:

1. Chromosome 11 Sequencing DataBases

   YAC/STS coordinates database
   cosmid end sequence database
   YAC-cosmid coordinate database
   Primers (new STSs)
   Homology/Identities listed by match significance

   Chromosome 11 sequencing data (complete cosmid/PAC sequences)
   11p15 project, 11p12 project
   WWW http://mcdermott.swmed.edu/
   Genbank

2. Clone libraries

   chromosome 11 cosmid 5X, arrayed
   chromosome 11 YAC ?X, arrayed (T. Shows/N. Nowak, RP)
   chromosome 11 and 15 PAC set in preparation

   (can be made available on request to G. Evans)

3. Software

   Mermade driver software for 192 channel synthesizer
   Primer prediction software for primer directed walking
   SUMU Lab sample tracking software
   Robotics control software for Biomek
   Data Inspector software for sequence quality control

   WWW http://mcdermott.swmed.edu/

4. Hardware specifications and construction plans

   Prepper III miniprep robot
   Mermade 192 channel oligonucleotidesynthesizer
   Lab workstations
   TREC multigel controller

Lab workstation plans and ordering information

WWW http://mcdermott.swmed.edu/

Available via:

WWW http://mcdermott.swmed.edu/

Any conditions attached:

Data resources are made available within 6 months after generation.

Hardware and software are supplied without warranty and without support other than helpful hints when needed. Hardware specifications and plans are available to all non-commerical users.

RESOURCES AVAILABLE

Name of participant : Chris Fields

Nature of resources available (software, maps, clones etc.)

Chris Fields

GSDB (complete, genome-scale relational DB)
scheduled for operational mid-summer

GSDB "Annotator" multiplatform client
interface (view/edit) available free
mid summer

Available via: http://www.hgvr.org

Any conditions attached: none

RESOURCES AVAILABLE

Name of participant : Richard A. Gibbs

Nature of resources available (software, maps, clones etc.)

- X chromosome mapped reagents - including binned cosmids (≥ 2,000) -
- Sequences, cosmids and Phe shotgun libraries from >1mb of human DNA from X, ch12 + ch17 available,
- matched cosmid/cDNA pairs available from X-chromosome, from C.C. Lee.

Available via:

All X chromosome + ch12 resources are described in their respective web pages.

Any conditions attached:

NO

RESOURCES AVAILABLE

Name of participant: Trevor Hutchins

Nature of resources available (software, maps, clones etc.)

>15,000 Human mapped STSs

~6,500 Mouse mapped SSRs

GRACE/BASS Gel analysis and basecalling software, UNIX based.

Primer Picking software (PRIMER 2.2)

Lab Base database system

Available via: http://www-genome.wi.mit.edu

Any conditions attached: None.

RESOURCES AVAILABLE

Name of participant : LaDeana Hillier

Nature of resources available (software, maps, clones etc.)

SOFTWARE : GETLANES (tracking gel images)
RETRAK (UNIX interface for editing lane tracking)
TPP (trace processing software)
PHRED (base calling)
PHRAP (sequence assembly)
FINISH (following shotgun completion, finish selects reads to contiguate & improve sequence quality)
DACE (implementation of a laboratory notebook tracking system in ACEDB)
other software tools are also available

Available via: HTTP://genome.wustl.edu/gschmfg.html

PHRED & PHRAP available: phg@u.washington.edu
ACEDB code available: ncbi.nlm.nih.gov /pub/repository/acedb

Any conditions attached:

retrak & tpp are still under intensive development.

RESOURCES AVAILABLE

Name of participant : PIETER DE JONG

Nature of resources available (software, maps, clones etc.)

Human PAC library (male donor, DNA from blood) 120 kb average insert (16-fold redundant, ~1200 384 well dishes)

Human PAC library (female donor, DNA from blood (~5-fold redundant) not yet arrayed; 150 kb insert.

Human BAC library : in progress, expect to deliver 10-fold redundant by May '96 and 20-fold by Summer '96.

Available via: PdJ, Roswell Park Cancer Institute

Any conditions attached:

- No secondary distribution of library, no problems to distribute individual clones (no ties attached).
- Cost-recovery of labor/plasticware/mailing costs for library replicates.

RESOURCES AVAILABLE

Name of participant: Dr. Hans Lehrach

Nature of resources available (software, maps, clones etc.)

The Resource Centre distributes high-density gridded filters of genomic libraries, cultures of individual library clones, or (in the future) PCR pools.

The table below gives details of those genomic libraries for which this service is now available, in the near future this will be supplemented with libraries from the I.M.A.G.E. consortium:

Library name	Description	Number
Cosmid (Human)		
L4/FS1	Chromosome 1 specific cosmid library	112
L4/FS6	Chromosome 6 specific cosmid library	109
L4/FS7	Chromosome 7 specific cosmid library	113
L4/FS11	Chromosome 11 specific cosmid library	107
L4/FS13	Chromosome 13 specific cosmid library	108
L4/FS17	Chromosome 17 specific cosmid library	105
L4/FS18	Chromosome 18 specific cosmid library	111
L4/FS21	Chromosome 21 specific cosmid library	102
L4/FS22	Chromosome 22 specific cosmid library	106
L4/FSC X/LA	Chromosome X specific cosmid library	101
L4/FSC X	Chromosome X specific cosmid library	104
Cosmid (other)		
L4/S.Pombe	S.pombe specific cosmid library	60
L4/B/S.Pombe	S.pombe specific cosmid library	61
Fugu-Cosmid	Fugu DNA partial cut with MboI in Lawrist4 and DH10B	66
P1		
P1 Human	Total Genomic P1 Human Library	700
MP1 Mouse P1 library	Total Genomic Mouse C57/Black6 P1 Library	703
pomP1	Schizosaccharomyces pombe (wt 972 h-) P1 library	705
PAC		
Human PAC	Human PAC library brought by Peter de Jong	704

RESOURCES AVAILABLE (Continued)

Name of participant: Dr. Hans Lehrach

Nature of resources available (software, maps, clones etc.)

(Continued from previous page)

Library name	Description	Number
YAC (Human)		
4X YAC	Human YAC library	900
4Y YAC	Human YAC library	901
CEPH YAC	Human CEPH YAC library	904
LSXY	Human YAC library	912
C3H YAC	Mouse YAC library	902
YAC (other)		
St.Marys Mouse YAC RAD52	Mouse YAC library from female C57BL/10 in host strain which is recombination deficient due to mutation in RAD52	909
C57 YAC	Mouse YAC library	905
Whitehead Mouse YAC I	Large insert Mouse YAC library constructed at the Whitehead Institute for Biomedical Research/MIT Center for Genome Research	910
pomYAC	Schizosaccharomyces pombe (wt 972 h-) YAC library	913
ICRF Pig YAC	Pig YAC library	907
LMUB Pig YAC	Pig YAC library from Lymphocytes (~300KB average inserts) from Ludwig Maximillian Univ.Muenchen	911
cDNA (Human)		
Human fetal brain cDNA	Human foetal brain cDNA made from 17 week embryo polyA+RNA	507
HFL cDNA	cDNA using dT primed polyA+ purified RNA from 21 weeks old human fetal liver	512
HTE cDNA	cDNA using dT primed polyA+ purified RNA from 21 weeks old human fetal thymus	508
HPO cDNA	cDNA from 21 weeks human foetal lung, poly dT primed, directionally cloned, excise enzyme MluI	515
cDNA (other)		
MBR cDNA	Mouse adult brain cDNA,synth: oligo dT primed,directionally cloned; cloning site: NotI/SalI; 1.5kb average insert size	510

RESOURCES AVAILABLE (continued)

Name of participant: Dr. Hans Lehrach

Nature of resources available (software, maps, clones etc.)

(see previous pages)

Available via:

The Resource Centre/Primary Database of the German Human Genome Project,
Max-Planck-Institut für Molekulare Genetik,
(Abteilung Lehrach),
Ihnestraße 73,
14195 Berlin (Dahlem)
GERMANY

Tel: +49 30 8413 1627

Fax: +49 30 8413 1395

WWW: http://rzpd.rz-berlin.mpg.de/

Any conditions attached:

Distribution of these resources will be free of charge to all participants in the German Human Genome Project, otherwise charges will be made to cover manufacturing expenses and postage costs.

In the case of some libraries additional conditions governing usage and distribution have been imposed by the owners.

RESOURCES AVAILABLE

Name of participant : DAVID J. LIPMAN

Nature of resources available (software, maps, clones etc.)

Databases ≠ Software
see : http://www.ncbi.nlm.nih.gov

Available via: WWW, FTP, CDROM

Any conditions attached: None

RESOURCES AVAILABLE

Name of participant:

Dr. Robert K. Moyzis
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Ph: 505-667-3912
FAX: 505-667-2891
email: moyzis@telomere.lanl.gov

Nature of resources available (software, maps, clones, etc.)

A) Complete digest libraries for each human chromosome
B) Partial digest phage and cosmid libraries for approximately half of the human karyotype (phage: 4, 5, 6, 8, 11, 13, 16, 17, X; cosmid: 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, X, Y)
C) YAC libraries for human chromosomes 9, 12, 16 and 21
D) M13/STS libraries (can be constructed for any human chromosome)
E) High-resolution YAC/STS/cosmid maps of human chromosomes 5 and 16

Available via:

A) American Type Culture Collection
B) Request from Los Alamos. Will also be available from commercial sources
C) Request from Los Alamos
D) Collaboration with Los Alamos
E) htpp://www-ls.lanl.gov; GDB and GSDB; request materials from Los Alamos

Any conditions attached:

A) Small fee; agreement to acknowledge Los Alamos in publications
B) Must sign Material Transfer Agreement with University of California limiting use to scientific purposes, limiting further distribution and agreeing to a limited collaboration with Los Alamos investigators
C) Collaboration with Los Alamos
D) Collaboration with Los Alamos
E) Sequencing coordinated with Los Alamos

RESOURCES AVAILABLE

Name of participant : Richard Myers + David Cox

Nature of resources available (software, maps, clones etc.)

- two panels of whole genome radiation hybrid DNAs
  (Stanford G3 panel - 400 Kb resolution)
  (Stanford TNG panel - 100 Kb resolution)
  available from Research Genetics
- map positions of 7300 STSs on the G3 radiation hybrids
- an email server allowing anonymous STS radiation hybrid scores to be integrated ~~on the~~ with our mapping data on the G3 hybrids

Available via: http://www-shgc.stanford.edu

Any conditions attached:

- none

RESOURCES AVAILABLE

Name of participant : Bruce Roe

Nature of resources available (software, maps, clones etc.)

Laboratory Protocols
Cosmid, P1 and BAC sequence data (In progress)

Available via: HTTP://dna1.chem.uoknor.edu

Any conditions attached:

Let us know if you find something cool that we missed

RESOURCES AVAILABLE

Name of participant : Melvin I. Simon

Nature of resources available (software, maps, clones etc.)

1. Mouse 129ESCell - BAC Library 235,000 clones (~10X coverage)
2. Human Fibroblast - BAC Library B 70,000 clones (~3X coverage)
3. Human Sperm BAC Library C 75,000 clones (~3x coverage)
4. Human Primary Fib. BAC Library A 100,000 clones (~4x coverage)
5. Human Sperm BAC Library D 75,000 clones
6. 619 - Ch 22 Specific Mapped BAC clones

Available via:

1, 2 and 3 Now Available - Research Genetics Inc (Huntsville Ala)
5 & 6 Available - Research Genetics Inc (April 1996)
4 - Available for screening via Hiroaki Shizuya - Biology Division Caltech - PASADENA FAX - (818) 796-7066 Also see: http://www.tree.caltech.edu

Any conditions attached:

No conditions or restrictions are attached to the material.

RESOURCES AVAILABLE

Name of participant :

Jim Weber

Nature of resources available (software, maps, clones etc.)

Crude, but comprehensive human linkage maps

STRP information

Methods

Image analysis software

Construction information for water bath thermal cycler and some SCAFUD components

Sequence assembly simulation program (from Gene Myers at University of Arizona)

Available via:

Website: http://genetics.mfldclin.edu
Email: gene@cs.arizona.edu

Any conditions attached:

Software is not supported.

RESOURCES AVAILABLE

Name of participant :

Jean Weissenbach

Nature of resources available (software, maps, clones etc.)

The Généthon Human Linkage Map
(5,264 microsatellite markers)

Map + description of reagents
(sequences, primers, alleles, frequencies, etc.)

Available via:

http://www.genethon.fr

Any conditions attached:

freely available

To: Collins, fc23a @ nih.gov>" <fc23a@nih.gov (Francis) @ INTERNET
cc: (bcc: Francis Collins/DIR/NCHGR)
From: GuyerM @ odder.nchgr.nih.gov ("Guyer, Mark") @ INTERNET
Date: 03/12/96 04:07:00 PM
Subject: Bermuda report

Here's the summary that we gave to the staff:

Summary of program staff meeting -- 2/29/96

Report on the International Strategy Meeting on Human Genome Sequencing

Mark and Jane both have copies of the full agenda and attendance list, if anyone wants to see them. The major groups that were represented were: Sulston, Waterston, Lander, Myers, Venter, Ansorge, Carrano, Moyzis, Evans, Caskey, Chen, Gibbs, Hood, Lehrach, Rosenthal, Weissenbach, McCombie, Roe, Weber, Hattori, Simon, de Jong, as well as Lipman, Fields, Ashburner. There was more than one person from several of the groups, a total of 50 people altogether, including agency types from Wellcome, NCHGR, DOE, MRC, Germany, Japan, and HUGO.

The major topics discussed were:

the sequencing plans/strategies/accomplishments of each of the groups;
the sequencing resources each group has and will make available (a list of these should have been distributed to staff);
data release;
data quality;
coordination among large sequencing groups.

Some of the key conclusions were:

most sequencing groups seem to be converging on a general strategy of using BACs selected by STS screening and a combination of shotgun and directed sequencing strategies; other strategies, including BAC end sequencing across the genome, shotgun sequencing of the entire genome, and sample sequencing across a complete chromosome were discussed, but in none of these cases was there group consensus that the strategy was superior to the generally-accepted paradigm;

data should be released regularly and very quickly from large-scale sequencing
projects, perhaps as frequently as daily but maybe weekly would do; this refers to preliminary data (i.e. contigs > 1 kb, not finished to database

submission quality) which would be put up locally automatically; it was also agreed that finished, annotated sequence would be immediately submitted to databases;

the attendees unanimously agreed to the following statement: for primary genomic sequence data from large-scale sequencing projects, the aim is to have all sequence freely available and in the public domain for both research and development, in order to maximize its benefit to society. This was intended to mean that the primary producers of the sequence from the Human Genome Project would not attempt to patent the sequence they generate. This statement was understood to be the sense of the attendees and that different organizations/agencies/countries might be under different constraints that might or might not allow them to adopt this as policy. The agencies were, however, urged to foster such policies.

data quality issues -- representation of data quality is becoming possible and
should be reported along with sequence data, particularly in the case of preliminary sequence; the group seemed to be moving toward agreement that the goal is 99.99% accuracy

International Coordination:

As a first step, each of the groups present discussed its goals for the next few years:

Seattle: 25-30 Mb in the next 3 years; primarily centered around the T cell alpha and delta regions on chromosome 7; also the human and mouse MHC regions

TIGR/Cal Tech: 30 Mb in 3 years on chromosome 16p

German consortium (administered by A. Rosenthal at Jena): 30 Mb of chromosome 21 (excluding the minimal Downs and PME regions); 1-2 megabase regions of X (Xq28, Xp11.2, the PAR1 region), 7, 11, 17

LANL: Moyzis original statement was that he intended to do one-pass sequencing across all of chromosome 16; by the time the meeting ended, he was reconsidering that and discussed producing finished sequence of regions of chromosome 16p not being pursued by Sanger or TIGR (a total of about 20 Mb) plus a region near the 5p telomere around the Cri du Chat locus

Whitehead: 105 Mb in 3 yr; human chromosome 17 and mouse syntenic regions plus a few random megabases here and there

Baylor: 30 Mb in 3 yr in Xq28, Xp22, 12p1.3

Oklahoma: I am funded to do 6 Mb in 3 years ; working on the region of 22q between NF2 and the centromere; this is being coordinated with Sanger and Wash U.

Japan: did not say, will report by correspondance after the meeting

Chen/ABD: 4 Mb in various regions on X (this is being coordinated through the X chromosome workshops)

LLNL: 50 Mb in 3-5 years on chromosome 19 and mouse syntenic regions

Wash U: 100 Mb in 3 years; regions on chromosomes 7, 22, and X to begin with

CSHL: 5 Mb from chromosomes 13 and 18

EMBL (Ansorge): cDNAs from chromosomes 21 and X (a total of 2 Mb)

Sanger: 150 Mb in 3 years (actually funded for 7 years to do 250 Mb); beginning with 22 and regions of X, then have all of 6 and 20 targeted, followed by chromosome 1

Marshfield: whole genomic shotgun sequencing will be pursued, hoping for a level of 2-4 Mb (raw?) per year

Dallas: 11p15.5, 11p12, 11q23 (2 mb in each region in 2 years); also pursuing sample sequencing across the whole chromosome.

Agencies: Each of the agencies made a short presentation about its plans. The gist of the information discussed was that the major funders for production human DNA sequencing will be NCHGR and the Wellcome Trust. There is a possibility that DOE and the German genome program will make some significant contributions. There was a brief allusion to developing French plans that could not be discussed in public at present. The European Union and the U.K. MRC will not be spending significant amounts of money on production sequencing. And the Japanese were silent.

There was general agreement that this had been quite a useful meeting and that it would be worthwhile reconvening (perhaps a smaller group) about a year from now.