"Bioinformatics Core Facilities: Training needs and other Issues"

GOBLET Workshop: Cape Town November 17, 2015

Fran Lewitter, Founding Director of Bioinformatics and Research Computing, Whitehead Institute for Biomedical Research, Cambridge MA USA

Bioinformatics and Research Computing (BaRC)

- **BaRC Home Page - http://jura.wi.mit.edu**

- 1 Director
- 3 Ph.D. Bioinformatics scientists
- 1 Masters
- 1 Graphics specialist

- 250 scientists in the institute; some "data heavy" labs have their own bioinformatics scientists

Articles of Interest

- Chang J. Core services: *Reward bioinformaticians*. Nature. 2015 Apr 9;520(7546):151-2. **PubMed PMID: 25855439**.

- Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, Schneider MV. *Bioinformatics curriculum guidelines: toward a definition of core competencies.* PLoS Comput Biol. 2014 Mar 6;10(3):e1003496. **PubMed PMID: 24603430.**

- Kallioniemi O, Wessels L, Valencia A. *On the organization of bioinformatics core services in biology-based research institutes*. Bioinformatics. 2011 May 15;27(10):1345. **PubMed PMID: 21402560**.

- Lewitter F, Rebhan M, Richter B, Sexton D. *The need for centralization of computational biology resources*. PLoS Comput Biol. 2009 Jun; 5(6):e1000372. **PubMed PMID: 19557184**.

- Lewitter F, Rebhan M. *Establishing a successful bioinformatics core facility team*. PLoS Comput Biol. 2009 Jun;5(6):e1000368. **PubMed PMID: 19557124**.

Core vs Embed

From: Kallioniemi O, Wessels L, Valencia A. Bioinformatics 2011

"In conclusion, we can state that we favor a model where a small core group of bioinformaticians provide transparently organized support on more general, institute-wide research problems, while the majority of the bioinformaticians are embedded in the research groups. Embedding ensures more direct interaction between the biologist and the bioinformatician, providing both researchers with a sense of ownership of the project. Not only does this elevate the skills level of both parties, but it also greatly enhances productivity."

Core vs Embed

From: Lewitter F, Rebhan M, Richter B, Sexton D. PLOS CB 2009.

- "...Below, we highlight some of the most important reasons we see for centralizing these resources…."
  - Provide infrastructure
  - Staffing issues
  - Cost/benefit

Who we train

High School Computer Science students (6)

High School Computer Science teacher sabbatical (1)

Undergrads – Biology, Math (3)

Bioinformatics Masters students (3)

Biology grad students (100s)

Biology postdocs (100s)

Hiring For a Core Facility

- Biologist or Computer professional?
- Bachelor level – least frequent hire
  - <u>Need</u> technical computer and good communication skills
  - <u>Lack</u> time and project management skills
- Master level
  - <u>Need</u> problem solving and interpretive skills; work independently; highly motivated
  - <u>Lack</u> experience in analysis of real biological data
- Ph.D. level
  - <u>Need</u> communication and management skills and ability to help others
  - <u>Lack</u> management skills, ability to synthesize information

Preliminary Conclusions from directors of core facilities (PubMed PMID: 24603430)

- Everyone wants smart, motivated people who can think critically and have deep domain knowledge.
- How people continue learning while working:
  - 29 of 29 respondents
    - Learn from group members
    - Self teach
  - 17 of 29 respondents
    - Take university courses
  - 6 of 29 respondents
    - Continuing education courses
  - Other
    - 9 say conference/workshops; 1 says vendors; 2 say collaboration; 1 says "they teach, they learn"

Main Issues Discussed

- Who to hire (biologist or computer scientist?)

- Core Facility vs. Embed in Lab?

- What skills are needed?

- How to keep people current?

Survey of Directors of Bioinformatics Core Facilities
10 of 27 respondents hire at the Bachelors level

BA Skills Needed		BA Skills Lacking	
General		**General**	
	communication		flexibility
	consulting skill		follow directions
	flexibility		keep good notes
	follow direction		manage multiple project
	keep good notes		project management
	passion		supervisory skills
	prob solv prog	2	time management
	smart		
3	work independently		

Domain Specific		Domain Specific	
	databases	2	biology
	experience distributed systems	2	statistics
	numeracy		tech competency
4	programming		
	soft eng		
	sys admin		
	tech lead		
	unix		

Numbers indicate # of responses, if greater than 1

Survey of Directors of Bioinformatics Core Facilities

22 of 27 respondents hire at the Masters level

	MA Skills Needed		MA Skills Lacking
	General		**General**
	able to apply knowledge		abstraction
	abstraction		breadth/experience
	generalization		generalization
	interpretive skills		critical thinking
	synthetic ability	2	prob solving
	problem solv		synthetic abil
	troubleshoot		troubleshoot

	MA Skills Needed		MA Skills Lacking
	assist/instruct users		independence
2	communication		job sched
	conduct experiments		communication
	curiosity		work in group
	independence		social skills
	planning to stay		
2	self motivated		
	very smart		

	Domain Specific		**Domain Specific**
2	advanced skill set in 1 area		bio or programming
4	bioinfo	4	biology
8	biology		structural bio
4	statistics		systems bio

	data analysis		analysis experience
	experience with seq data		aware of pub dom tools/resources
2	lab experience		exper hi seq platforms
	micro ar/gene expr		lit searches

	databases		database dev
14	programming		linux admin
2	R		PC/Mac supp
	sys admin	3	programming
	technical expertise		software documentation
2	unix	2	SQL
		6	Statistics
		2	technical expertise

Numbers indicate # of responses, if greater than 1

Ph.D. Skills — General

#	Ph.D. Skills Needed — General	#	Ph.D. Skills Lacking — General
2	ability to synthesize		critical thinking
2	critical thinking		dual experience
3	interperative skills		experience
	pick up skills on fly		synthetic abil
2	problem solve		transl theory to practice

#	Ph.D. Skills Needed — General	#	Ph.D. Skills Lacking — General
	adaptability/passion	2	abil to complete proj
5	communications		altruistic behavior
2	independent	4	communications
	learn+help others		empathy
	manage/trouble workflows		focus & reporting
2	management skills		independent research
	people skills		manage multiple projects
	social skills		manage/trouble workflows
	strength+breadth		project management
	team leader qualities		social skills
			working in service environ

Ph.D. Skills — Domain Specific

#	Ph.D. Skills Needed — Domain Specific	#	Ph.D. Skills Lacking — Domain Specific
10	bioinfo knowledge/experience		aware of pub dom tools/resources
	interdisci research		biology
2	next gen		comp bio knowlegde
	formulate new research		genomics
	genomics		interdisci resear
	participate in research		real bioinfo experience
	prior industry exp		systems bio
	prior lab experience		x-train biol & programming
	strong bioinfo		
	tech exper		

#	Ph.D. Skills Needed — Domain Specific	#	Ph.D. Skills Lacking — Domain Specific
	algorithm development		analysis experience
	data analy methods		apprec data qual
	data curation		data curation experience
	data mining		develop products
	experiment planning		formulate new research
	databases		heterog comp env experi
	expression array analysis		snp analysis
	imple new protoc/train		
	instrumentation		

#	Ph.D. Skills Needed — Domain Specific	#	Ph.D. Skills Lacking — Domain Specific
	linux		instrumentation
	pathway analysis	2	linux admin
8	programming/scripting/Perl	3	programming
6	statistics	2	robust coding skills
	stats/bioinfo data anal		software framework planning
			sql
		3	statistics
			tech exper
			unix

Survey of Directors of Bioinformatics Core Facilities

25 of 27 respondents hire at the Ph.D. level

Numbers indicate # of responses, if greater than 1