

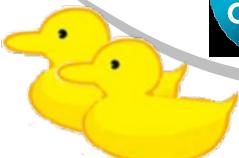
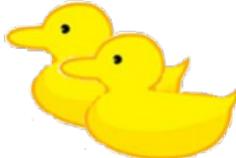
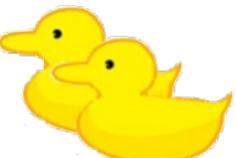
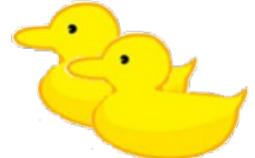
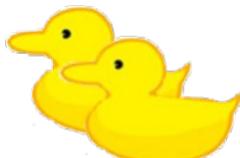
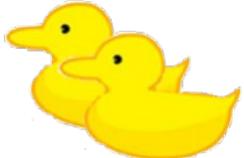


Close-Kin Mark-Recapture for fisheries and bycatch



Mark Bravington: Aug 2022

O&A
www.csiro.au



Acknowledgements



People (at CSIRO except as noted)	Organizations (Aus. except as noted)
Jorden Aulich	RITF Benoa: Indonesia
Shane Baylis	DoE (NERP/NESP)
Russ Bradford	FRDC
Barry Bruce	AFMA
Paul Conn: NOAA	NSW DPI
Scott Cooper	NT Fisheries
Campbell Davies	DaRT PL
Bill de la Mare: AAD	CCSBT: global
Eric Anderson: NOAA	ICCAT: global
Paige Eveson	Charles Darwin Uni
Jess Farley	Flinders Uni
Pierre Feutry	C-STAR NOAA/UCSC: USA
Peter Grewe	Bergen Uni: Norway
Rasanthi Gunasekara	IFREMER, Nantes: France
Peta Hill	Alaska DFG / NMML Seattle
Laura Tremblay-Boyer	Lenfest Ocean Program: global
Hans Skaug	
Tore Schweder	

CKMR models/designs: fisheries 2022

1. Intro

- ideas / projects / scope / history
- kinship probs and critiquery
- putting them together--- “The Framework”

2. POP examples--- mammals

- can we break it?
- different sampling issues

3. Super-simple POPs

4. HSP examples--- mammal-like sharks

5. Fish time! POPs + HSPs

- what does CKMR *really* tell us?

6. Genetics in 3 nutshells

7. Spatial stuff & BYO species

8. Design

CKMR models/designs: fisheries 2022

1. Intro

- ideas / projects / scope / history
- kinship probs and critiquery
- putting them together

2. POP examples--- mammals

- can we break it?
- different sampling issues

3. Super-simple POPs

4. HSP examples--- mammal-like

5. Fish time! POPs + HSPs

- what does CKMR *really*

6. Genetics in 3 nutshells

7. Spatial stuff & BYO species

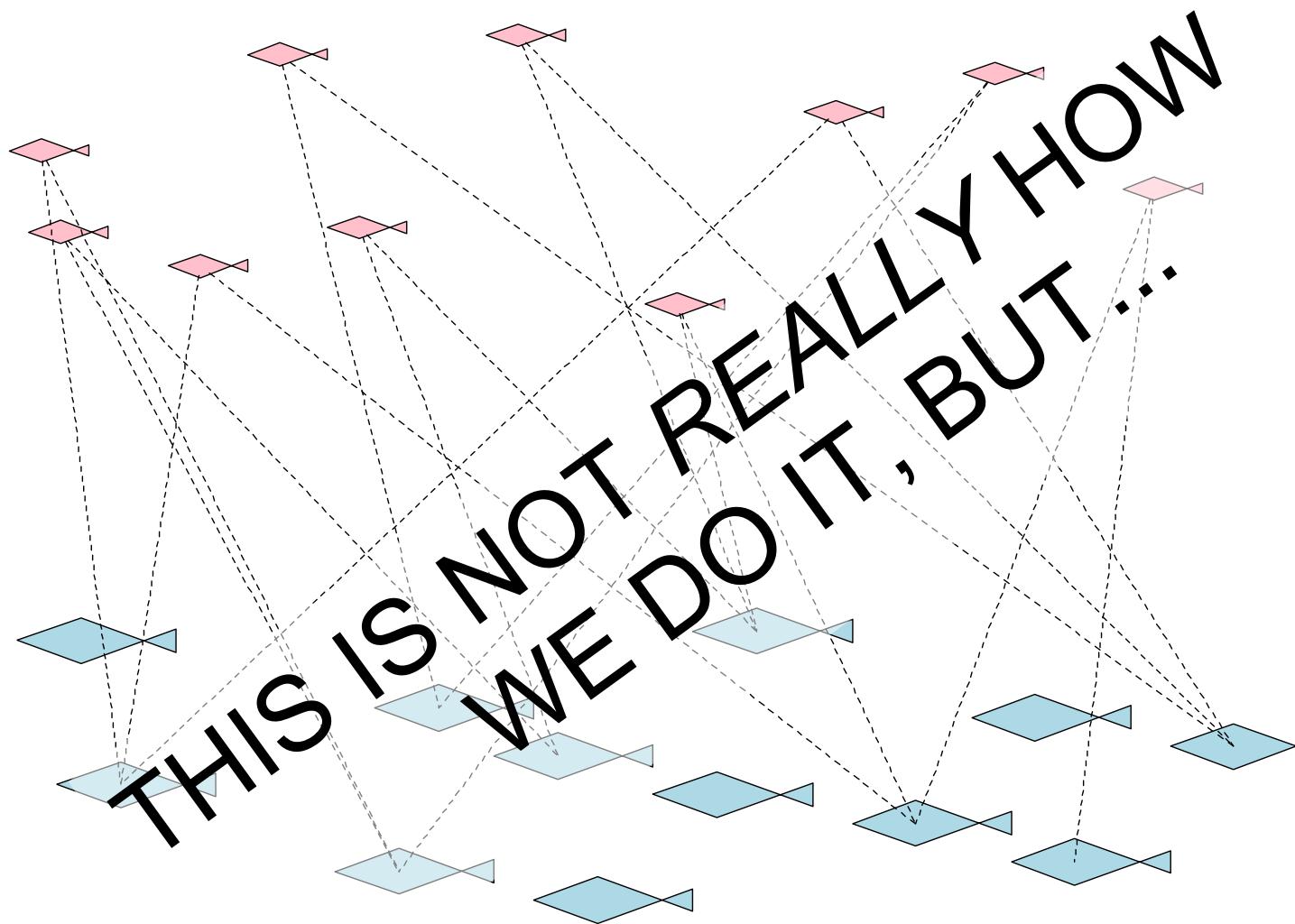
8. Design

By the end, you should...

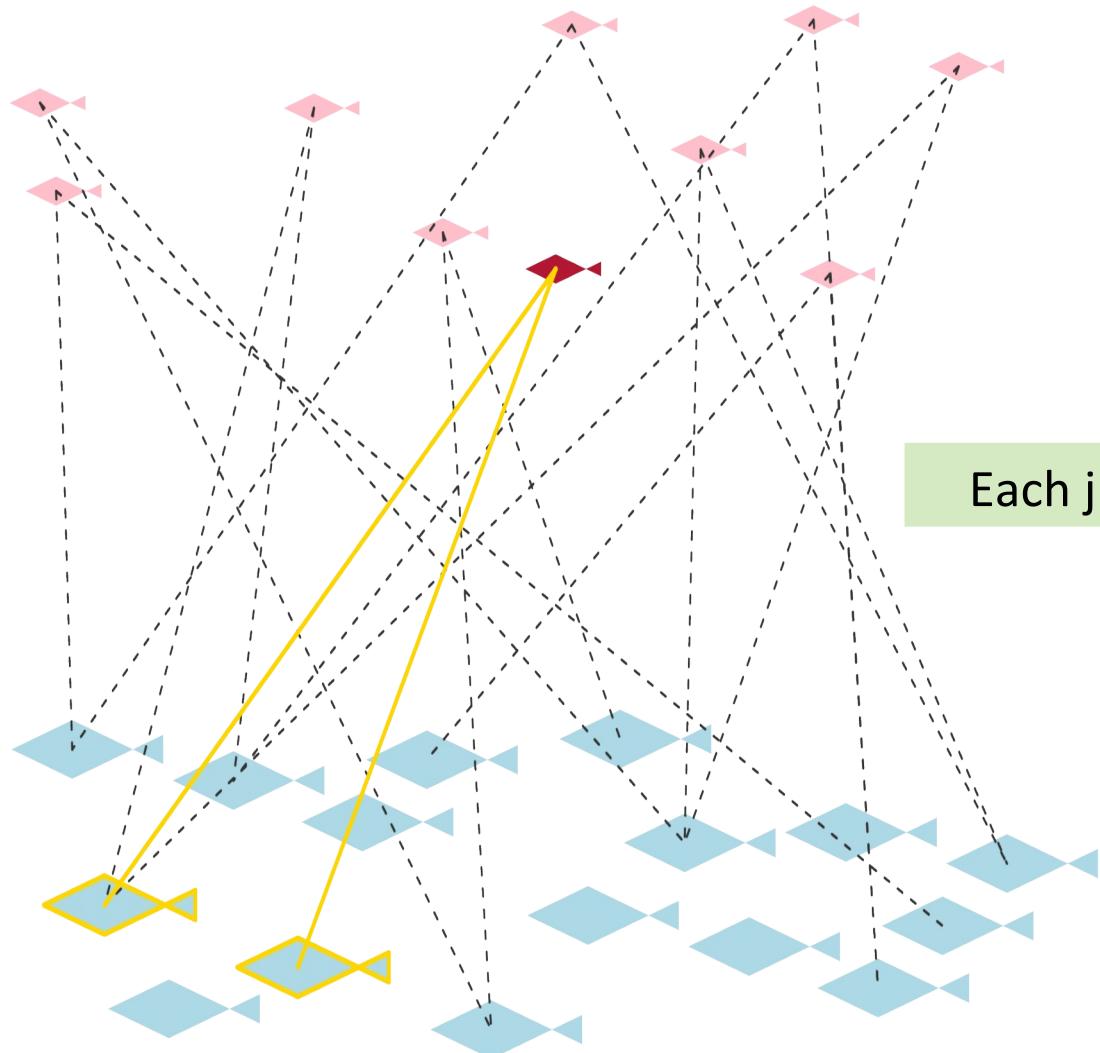
- understand how to build basic CKMR models
- understand how to critique other people's :)
- have more idea about genetics than you do now[*]
- have some idea of pitfalls to avoid
- have some idea how to go about CK Design
- know more cryptobiology than you do now



The \$1,000,000 cartoon

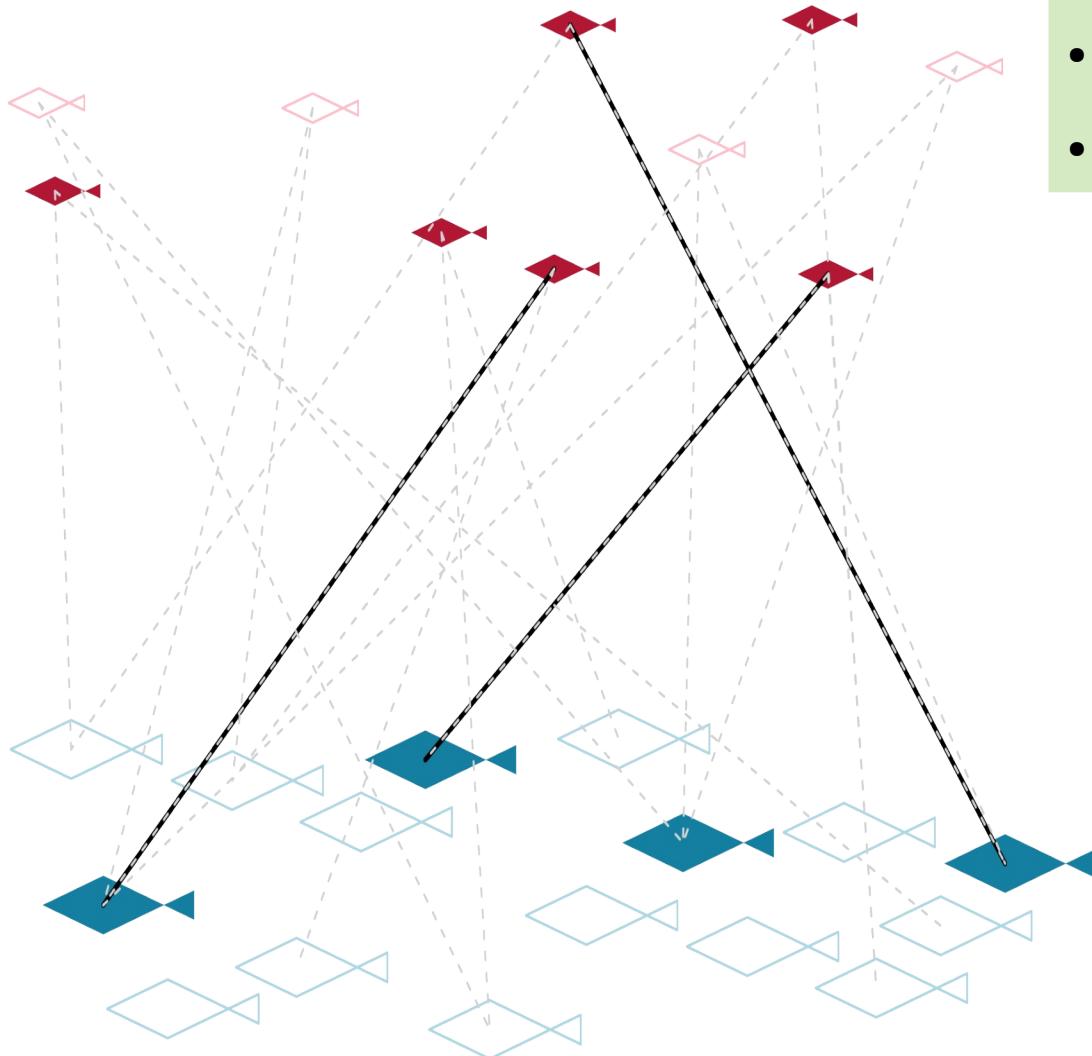


CKMR: the POP* cartoon



*POP: Parent - Offspring Pair

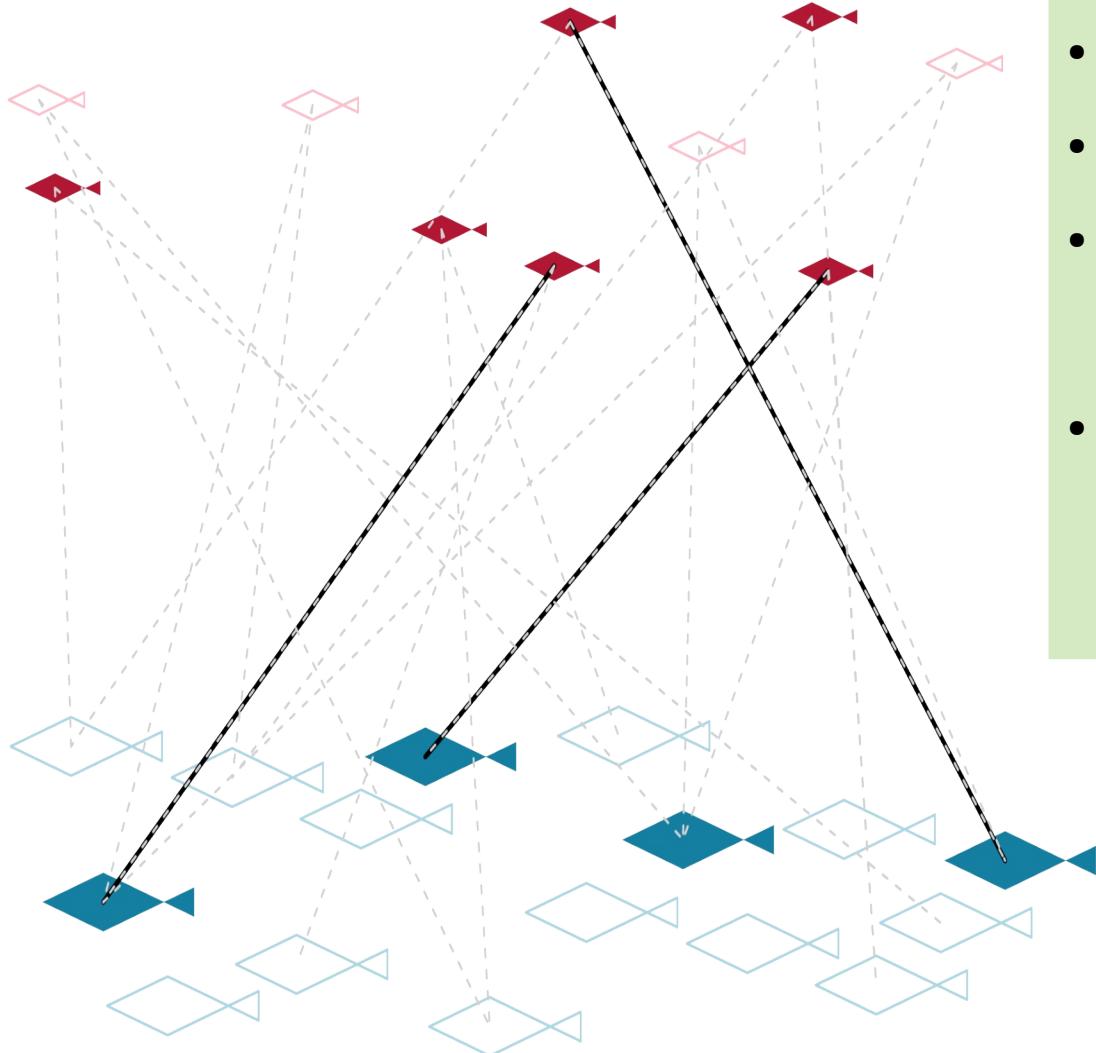
CKMR: the POP* cartoon



Sample adults and juvs

- ... genotype them ...
- ... look for POPs ("marks")

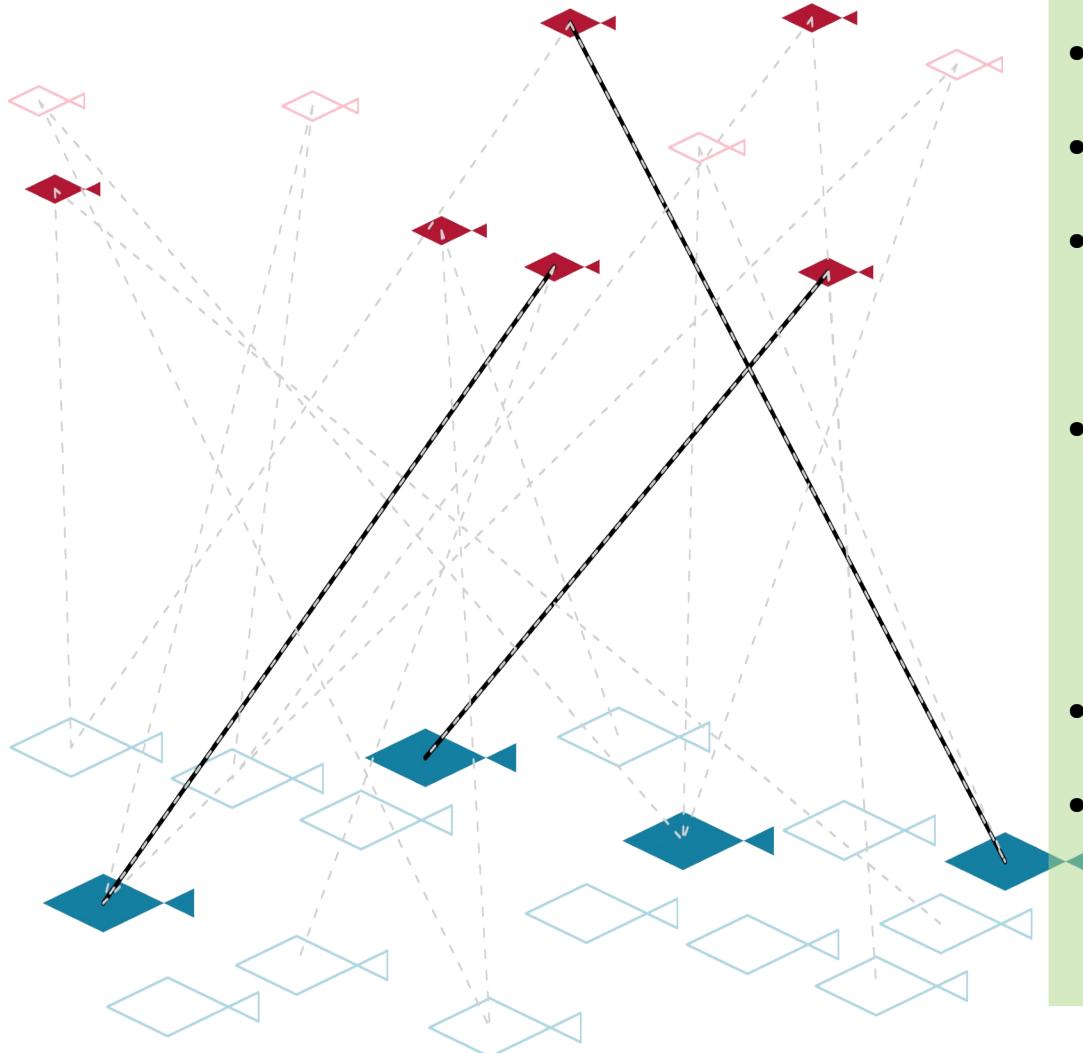
CKMR: the POP* cartoon



Sample adults and juves

- ... genotype them ...
- ... look for POPs ("marks")
- Each adult-juve comparison:
 $\Pr[\text{POP}] = 2/N_{\text{adult}}$
- Sample 6 juves & 4 adults
 - 24 pairwise comparisons
 - 3 POPs found

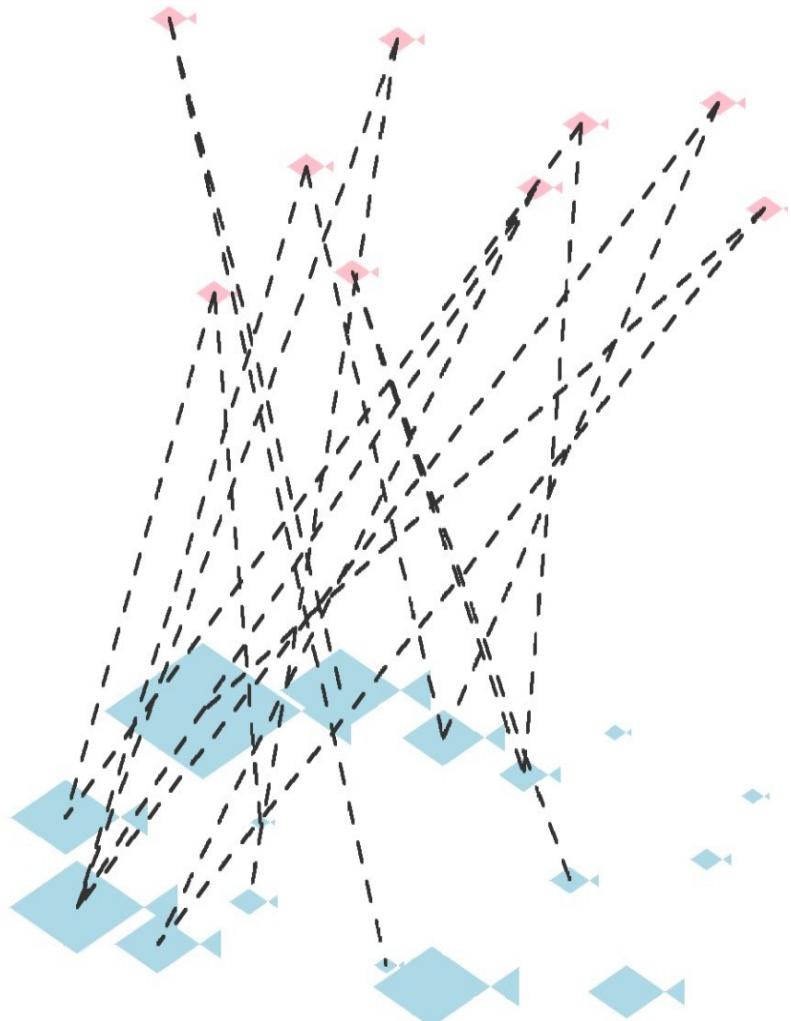
CKMR: the POP* cartoon



Sample adults and juves

- ... genotype them ...
- ... look for POPs ("marks")
- Each adult-juve comparison:
 $\Pr[\text{POP}] = 2/N_{\text{adult}}$
- Sample 6 juves & 4 adults
 - 24 pairwise comparisons
 - 3 POPs found
- $24 * 2/N_{\text{adult}}$ gives 3
- "MLE" of N_{adult} is 16
 - ... which happens to be the right answer. Lucky !

But what if...



- Bigger fish more fecund?
- ... more likely to be caught?
- Fish caught at different times?

Need a way to generalize
beyond “ $2/N$ ”

CKMR is...

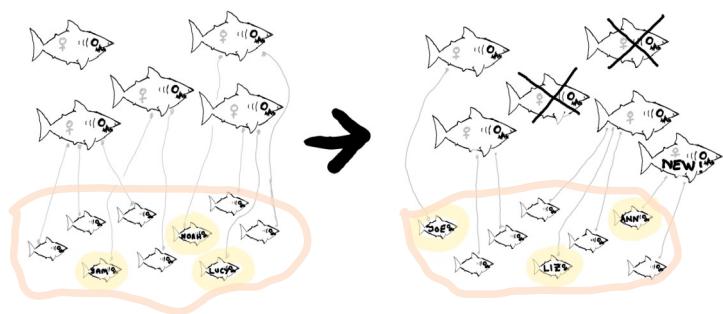
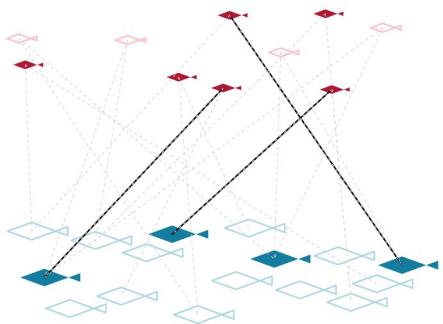
cross-cohort half-sibs

Parents are “marked” by their sampled offspring

Direct recapture (POPs)

and

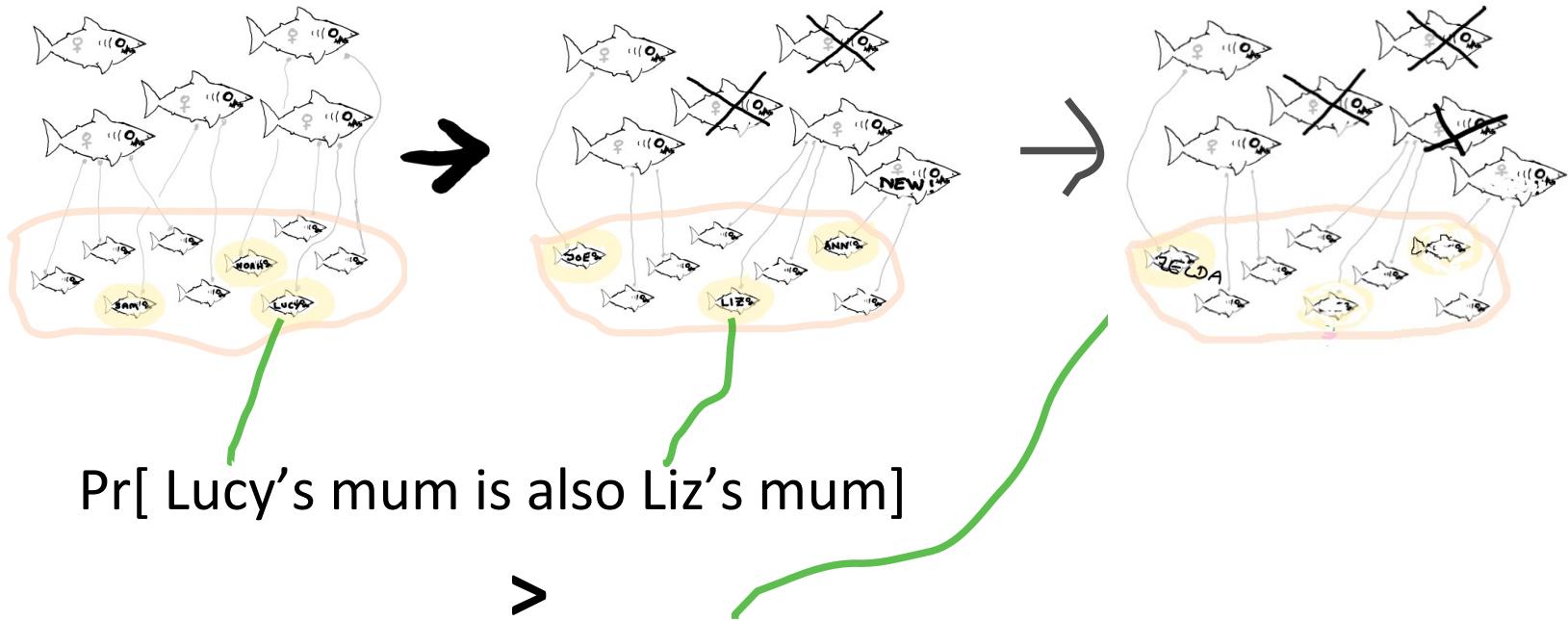
Indirect (XHSPs)



$$\Pr[\text{POP}] = O(1/N_{ad})$$

$$\Pr[\text{XHSP}] = O(1/N_{ad}) * \dots$$

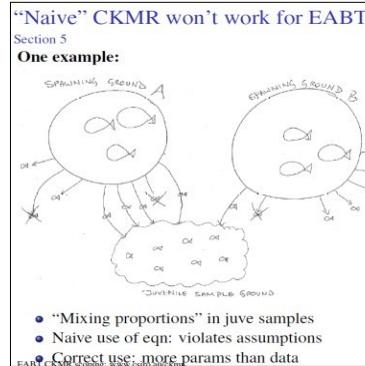
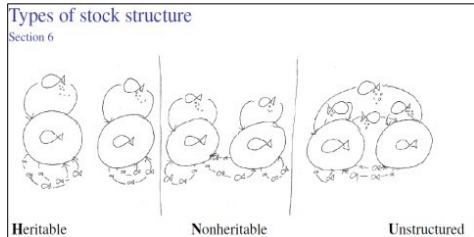
HSPs: abundance and mortality info



NB: Things can be a bit more complicated than with POPs

Stock structure in CKMR

- When stock structure exists, "naive" CKMR *sometimes* biased
- But, the kin-pairs themselves **tell you the real structure !**
 - even when no pop-gen signal
- *"Can we fix it? Yes we can!" [*]*
 - (Builder, B.; *CBeebies*; 2005)
 - ... if your sampling is adequate

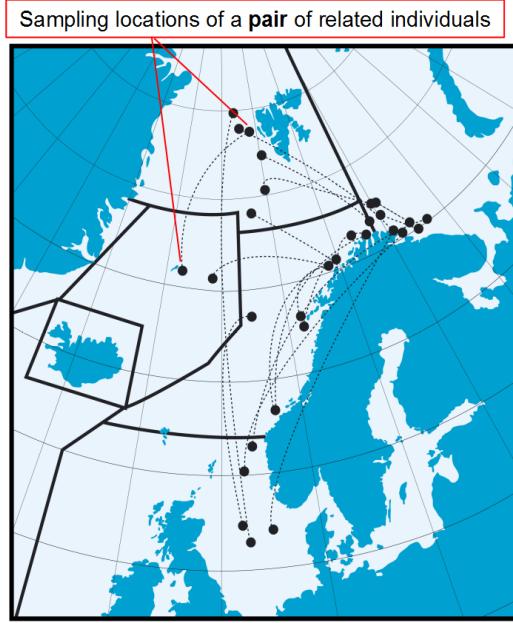


Norwegian minke whale DNA-register

- 10,000 sampled individuals
 - Sparse sampling!
- 10 microsat markers

Detection of relatives by forensic methods (Skaug et al, 2010, Mol. Ecol. Res.)

- n = 3300 (years 1997-2002)
- 21 pairs detected



Bob the Builder is a *notoriously* optimistic chap!

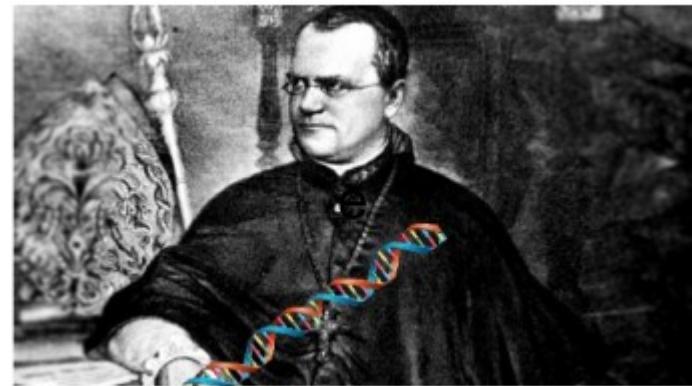
Sally the Scientist says “*Sometimes; it all depends...*”



CKMR core assumptions

1. Every animal was born with 1 mother and 1 father
2. You can reliably detect POPs and HSPs via genetics*

*Everything else is "just"
maths and logistics...*



*Or, at least *most* of them; see later

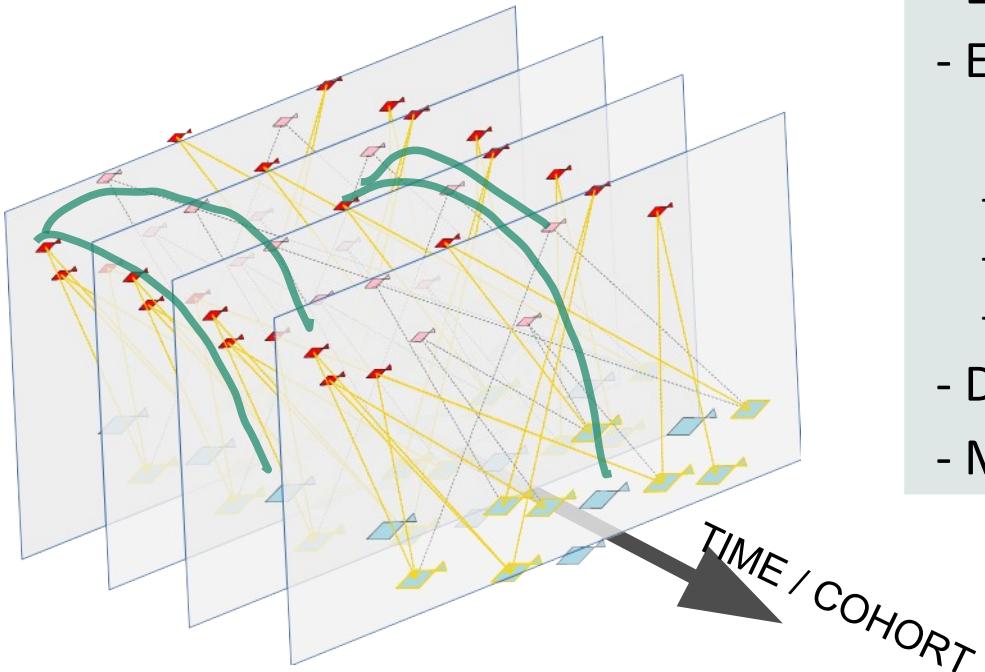
CKMR is...

Parents are “marked” by their sampled offspring

Direct recapture (POPs)

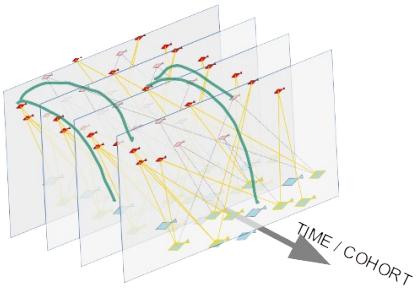
and

Indirect (XHSPs)



- Lots of comparisons
- Each sample used Xtuple times
 - ... with several roles:
 - potential Parent
 - potential Offspring
 - pot. Off as sib
- Different prob formulae
- More parameters than just “N”

Part of the art of CKMR is...



... deciding which comparisons *not* to use

- some affected by “nuisance” as well as “interest”
 - too much trouble to model
-
- Fine to leave some out, but don’t cheat!
 - i.e. don’t look at comp outcome *then* decide



“Executive”-level CK pitch

$$\begin{array}{lll} \text{POPs} & \Rightarrow & N \\ \text{HSPs} & \Rightarrow & Z \\ Z = F + M & = & C/N + M \\ M & = & Z - C/N \end{array}$$

- all just for Adults, of course
- N, C time-varying--- fit it all in a model, don't adhockerize
- do separately by sex
- caveat blah blah etc

For “non-executives”:
this is OK for Mammals but *a bit*
oversimplified for Fish
nevertheless still true for Fish

A page of (mostly) nice things

- Absolute **N ; m; stock structure** insights
- **Samplees already caught for you**
 - pretty cheap (\$█ per sample... *small* ppn of catch)
- Auto-back-dated; quick/good mixing
- “Let the fish come to you”:
 - don’t have to sample *every* fishery, *every* location
- Can’t cheat
- Conditioned on *what* samples you’ve got...
 - not *why* you’ve got them, nor how many
- Model *fish biology* not *human psychology*
- Minimally subjective
 - “automatic” assessment; easy MSE-testing
- Vague good news / precise bad news
- Precision not good enough? Just \$pend a bit more

- Direct info only about adults

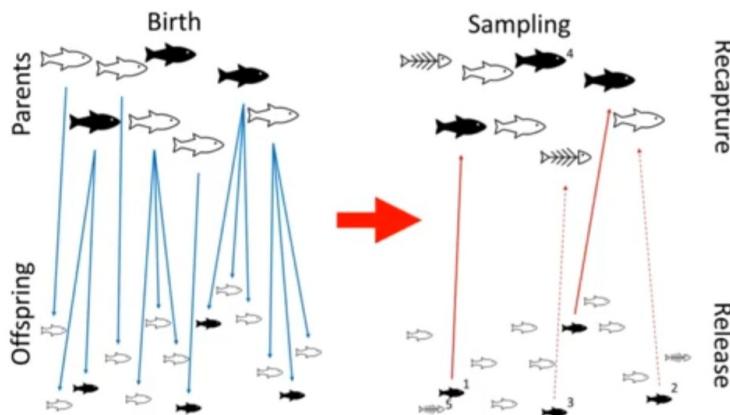


slide courtesy of Mark Maunder

A no BS guide to fishery stock assessment

The solution: Close-Kin Mark-Recapture

1. Stock structure
2. Adult natural mortality (survival)
3. Absolute spawner biomass estimates
 - Still need to estimate
 - Juvenile abundance
 - Juvenile natural mortality
 - Stock-recruitment relationship
 - Benefits
 - Can use dead fish so wider sampling coverage
 - No tag induced mortality
 - No tag loss
 - No miss reporting
 - Larval and juvenile dispersal distributes the tags spatially



Bravington, Mark V., Peter M. Grew, and Campbell R. Davies (2016). "Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture". In: Nature Communications 7. Skaug 2001. Allele-sharing methods for estimation of population size. Biometrics 57, 750-756.

Hillary et al. 2018. Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. Scientific reports 8, 2661-2661.

Bravington et al. 2016. Close-Kin Mark-Recapture. Statistical Science 31, 259-274.

all more-distant kinships. To estimate demographic parameters, the results of all pairwise comparisons are embedded in an extended mark-recapture (MR) framework, adapted to the biology and sampling arrangements of the species and study system in question.

The approach has several useful, indeed almost magical, properties:

Time travel:

marking happened back when the samples were born, which, depending on their ages, may date to before the study's conception. Thus, CKMR projects do not need to wait years for samples to "mix" or demographics to take effect; the samples' own lifespans are a window back in time. This can give us decades-long understanding of demographic parameters (e.g., growth rate), without the need for a decades-long study, depending on the lifespan of the study species.

Crystal balls:

learning about adults without ever seeing them, by finding their offspring as half-siblings. This is important for some species that cannot readily be sampled as adults, including large sharks.

Reincarnation:

recapturing an animal after it is dead, via its surviving offspring.

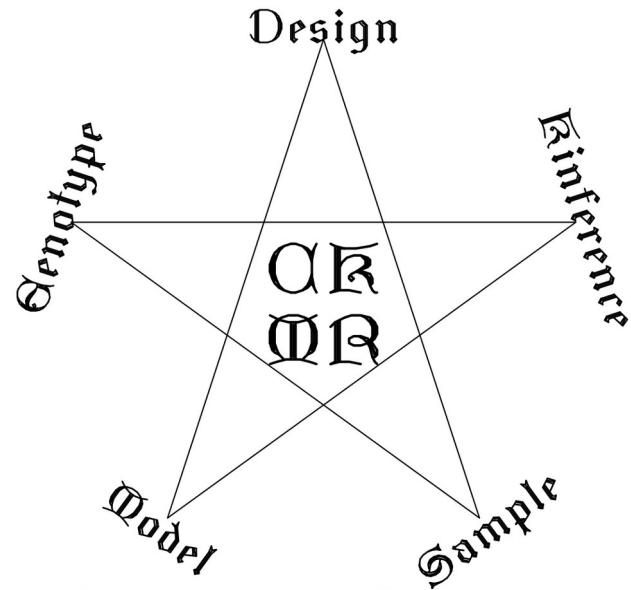
A key point is that samples can if necessary come *only* from dead animals; unlike conventional MR, there is no need for any live-release, which can be an expensive or impossible undertaking for many species. CKMR is therefore well suited to settings like fisheries and hunting, where large sample sizes can be obtained incidentally at little cost.

Of course, it is also fine to collect biopsy samples from live animals. CKMR is very useful for endangered species as well as exploited ones, both in its own right and in conjunction

CKMR is a 4-letter word...

easy

MAGICAL
RESULTS ...

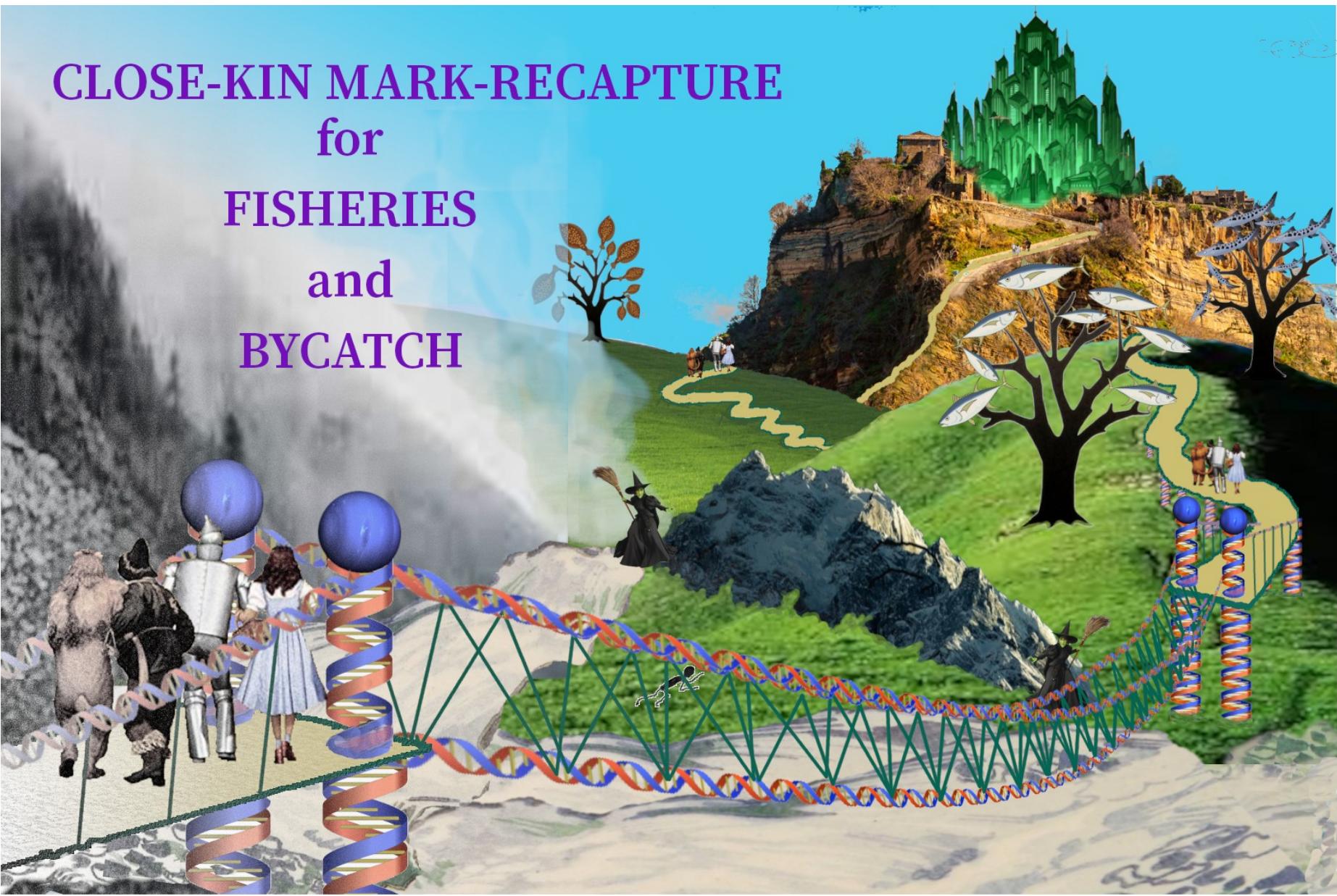


come from
CAREFUL
ENGINEERING !

... and assembling the RIGHT TEAM ...



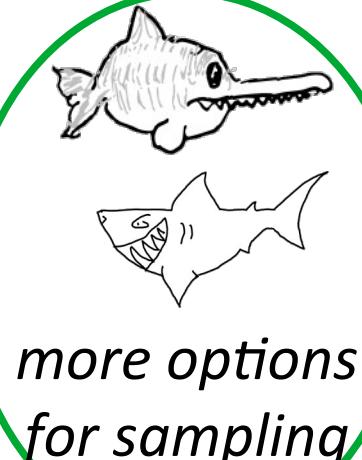
CLOSE-KIN MARK-RECAPTURE for FISHERIES and BYCATCH



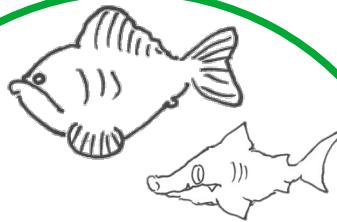
CLOSE-KIN TREE OF LIFE

simplified version!

adult fecundity
does not change



adults grow a lot



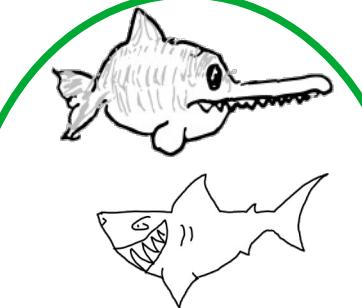
breed-and-die



CLOSE-KIN TREE OF LIFE

simplified version!

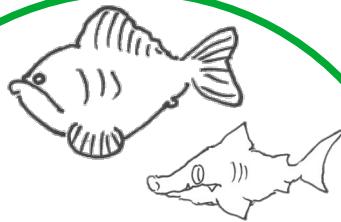
adult fecundity
does not change



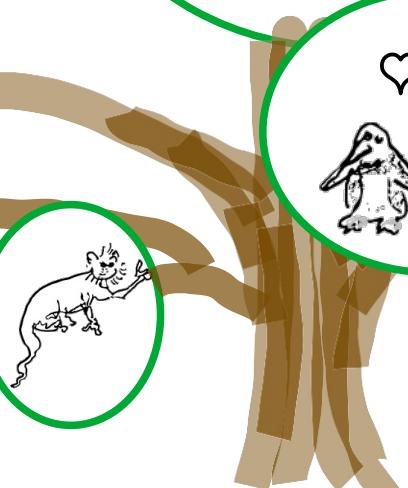
*more options
for sampling*



adults grow a lot



*demanding!
only one good
way to sample*



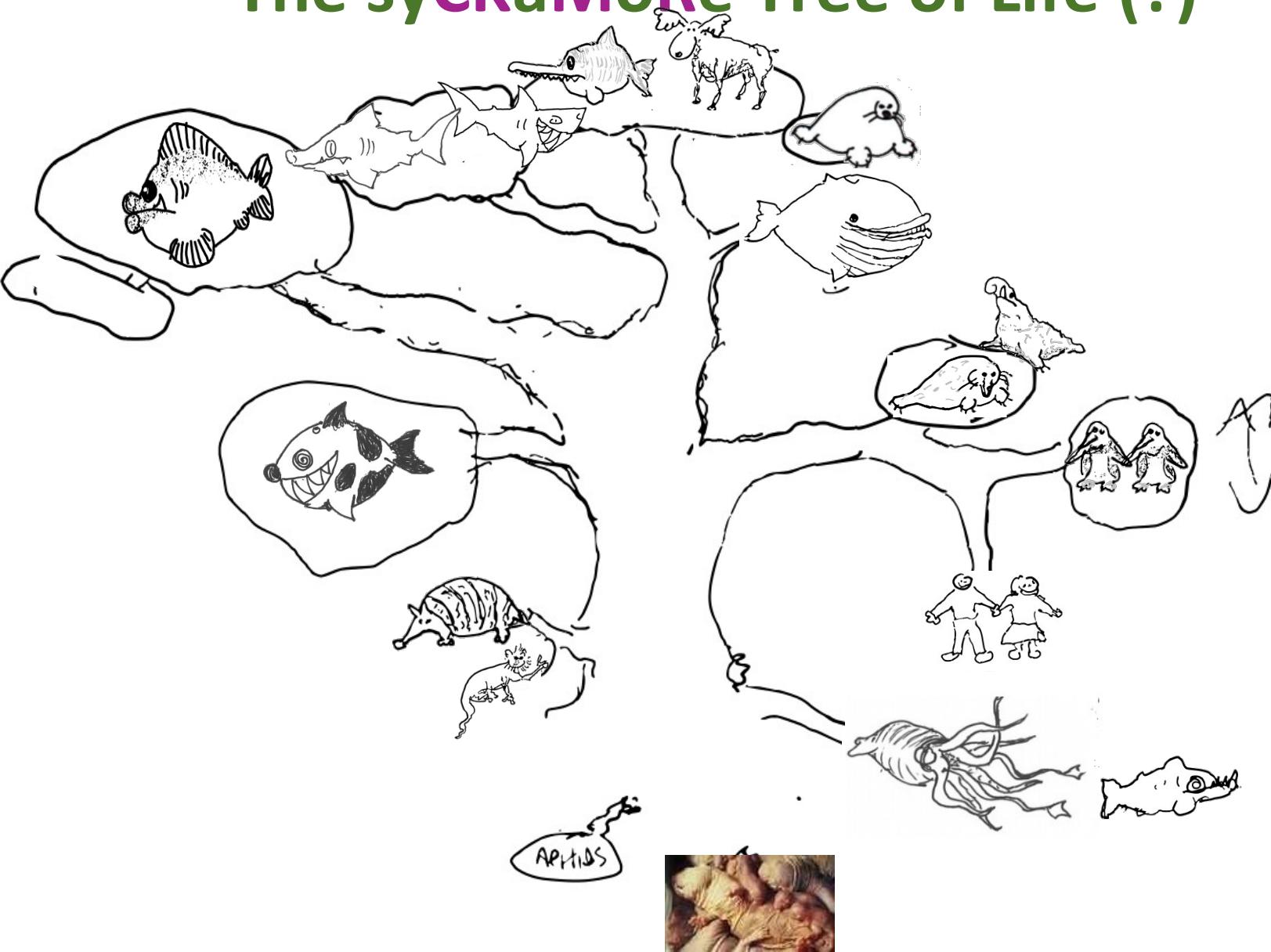
breed-and-die



*forget it
(except if...)*

ETC ... ETC

The syCKaMoRe Tree of Life (?)



Scope, sample size, etc (fish-oriented)

(Do this on whiteboard, unless there's been time to prepare – or find – a series of slides...)

Scope, sample size, etc (fish-oriented)

- Very short-lived, very long-lived: probably not. Otherwise:

For OK precision, need $\geq 50\text{--}100$ POPs/XHSPs, regardless of N

Say $B = \#samps$. Then $\#comps = \text{“const0”} * B^2$

Each comp has $\Pr[\text{kin}] = \text{“const1”} / N$

Expected #kin-pairs = $\text{“const2”} * B^2 / N = 100$

$\Rightarrow B = \text{“const3”} * \sqrt{N}$

But: $C = \text{“const4”} * N$

Bigger populations mean more samples but even more catch

$B/C = \text{“const5”} / \sqrt{N}$: **bigger is better value**

caveats, caveats...



Some myths, busted

“CKMR needs validation”



Some myths, busted

“CKMR needs validation”

That is NONSENSE! The basic assumptions are:

- everything had 1 mother and 1 father at birth;
- genetics lets you find POPs and HSPs

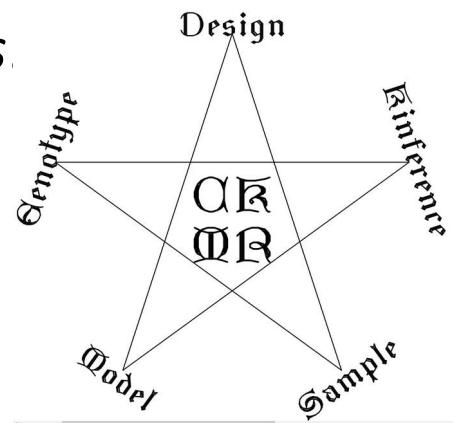
The rest follows by *ironclad laws of mathematics*.

That said:

*of course people could completely stuff up
any particular CKMR study*

So: *check each study carefully!* can't “validate” directly-- against what???

But that has nothing to do with “validating CKMR” in general



Some myths, busted

~~CKMR needs validation~~

“It’s got to be done exactly like SBT c. 2014”



Some myths, busted

~~CKMR needs validation~~

“It’s got to be done exactly like SBT c. 2014”

No it doesn’t! CKMR is fairly flexible.

EG ideally you *don’t* want to sample spawning grounds...
you *don’t* absolutely need adults-and-juves well-mixed tho’ it is best
you *do* need POPs and HSPs **for fish but not necessarily sharks**

- we had to use a special trick in SBT mk I POPs-only...
- ... which won’t work in other species...
- ... and turned out not to be fully valid ...
- ... when we did mk II POPs+HSPs. But it’s all good now !

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

“I've seen the cartoon. That $2/N$ formula will be biased if...”



Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

“I've seen the cartoon. That $2/N$ formula will be biased if...”

Of course it won't work! *You're not meant to use $2/N$ for real...*

CKMR theory tells you how to adapt the idea for real situations. Need a qualitative understanding of the biology and sampling, plus (a bit of) maths. So, do that instead !!! And then you will *avoid bias*.



Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That $2/N$ formula will be biased if...~~

“It’s expensive...”

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That $2/N$ formula will be biased if...~~

"It's expensive..."

It is NOT* ! You sample a *tiny fraction* of the catch that happens anyway; each one is cheap to genotype (<\$ █).

Much cheaper than dedicated surveys or normal MR if those even apply

Sure--- CPUE is free. Super! You totally get what you pay for there.

Plus, co\$t of mi\$management due to mi\$leading conventional data.

May be uneconomic for small pop'ns of low-unit-value short-lived spp.

* Unless you have to do something weird



Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That $2/N$ formula will be biased if...~~

~~It's expensive...~~

“It needs really careful attention-to-detail and a lot of know-how”

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That $2/N$ formula will be biased if...~~

~~It's expensive...~~

“It needs really careful attention-to-detail and a lot of know-how”

This one is **not** a myth! If you blunder into CKMR without a clear plan and without the right team, *things will end badly*. So: don't do that!



Epigenetic age (methylation-based, “DNAge”)

cite: Ben Mayne / Simon Jarman etc

To date, **worst CK problems from lousy age and/or length data**

- DNAge development ongoing at CSIRO and elsewhere
 - prelim results 4 spp: from decent to amazing !
- **Cheap!** Under \$ █ per sample and low setup cost
- A(nother) game-changer and useful beyond CKMR
- Needs calibration per species: research
 - tho', as of Aug 2022: can be done *in-model*
- All you'll need: **a quick jab**
 - length nice, but not essential
- More on Thu



Epigenetic age (methylation-based, “DNAge”)

cite: Ben Mayne / Simon Jarman etc

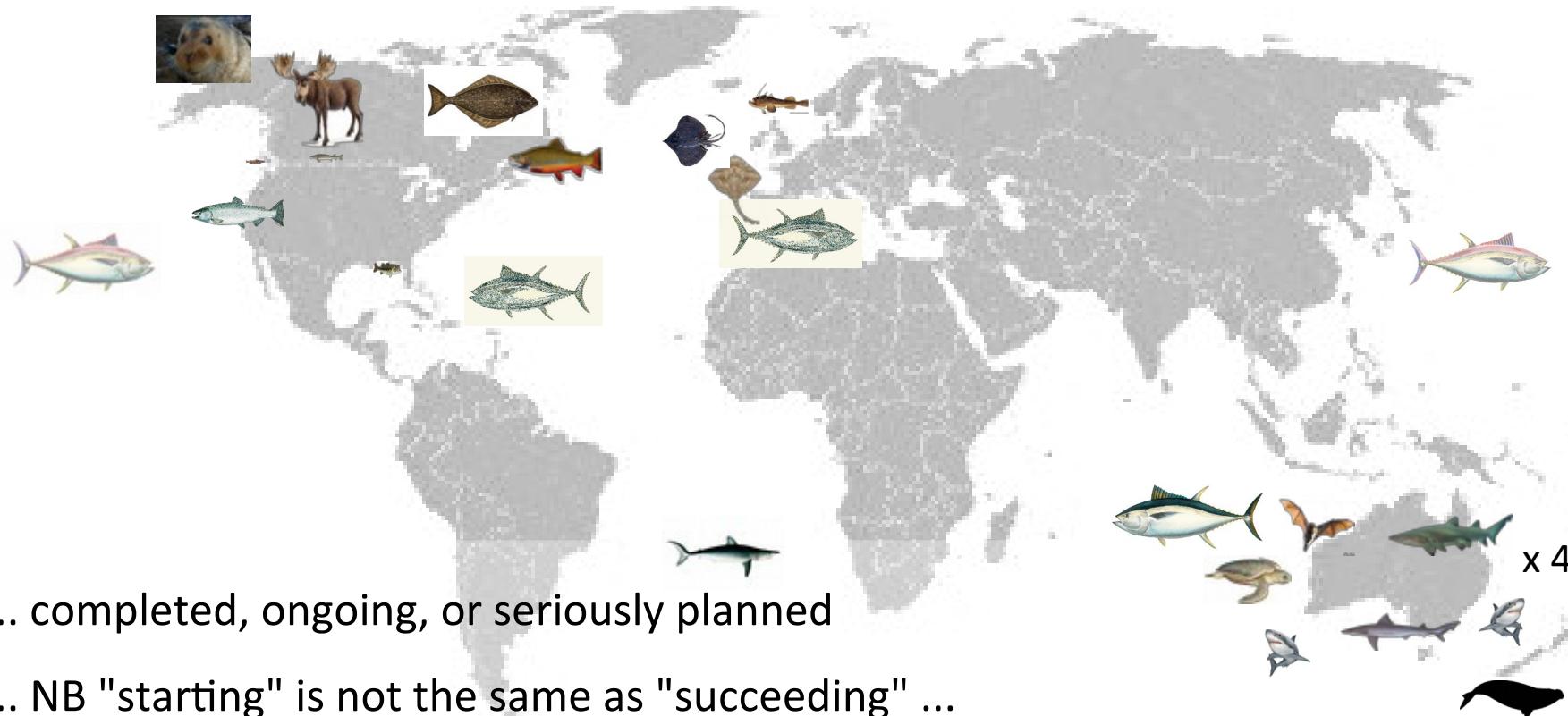
To date, **worst CK problems from lousy age and/or length data**

- DNAge devel ongoing at CSIRO and elsewhere
 - prelim results 4 spp: from decent to amazing !
- **Cheap!** Under \$ █ per sample and low setup cost
- A(nother) game-changer and useful beyond CKMR
- Needs calibration per species: research
 - tho', as of Aug 2022: can be done *in-model*
- All you'll need: **a quick jab**
 - length nice, but not essential

!!! WOW!!!



(some) CKMR projects c. 2022



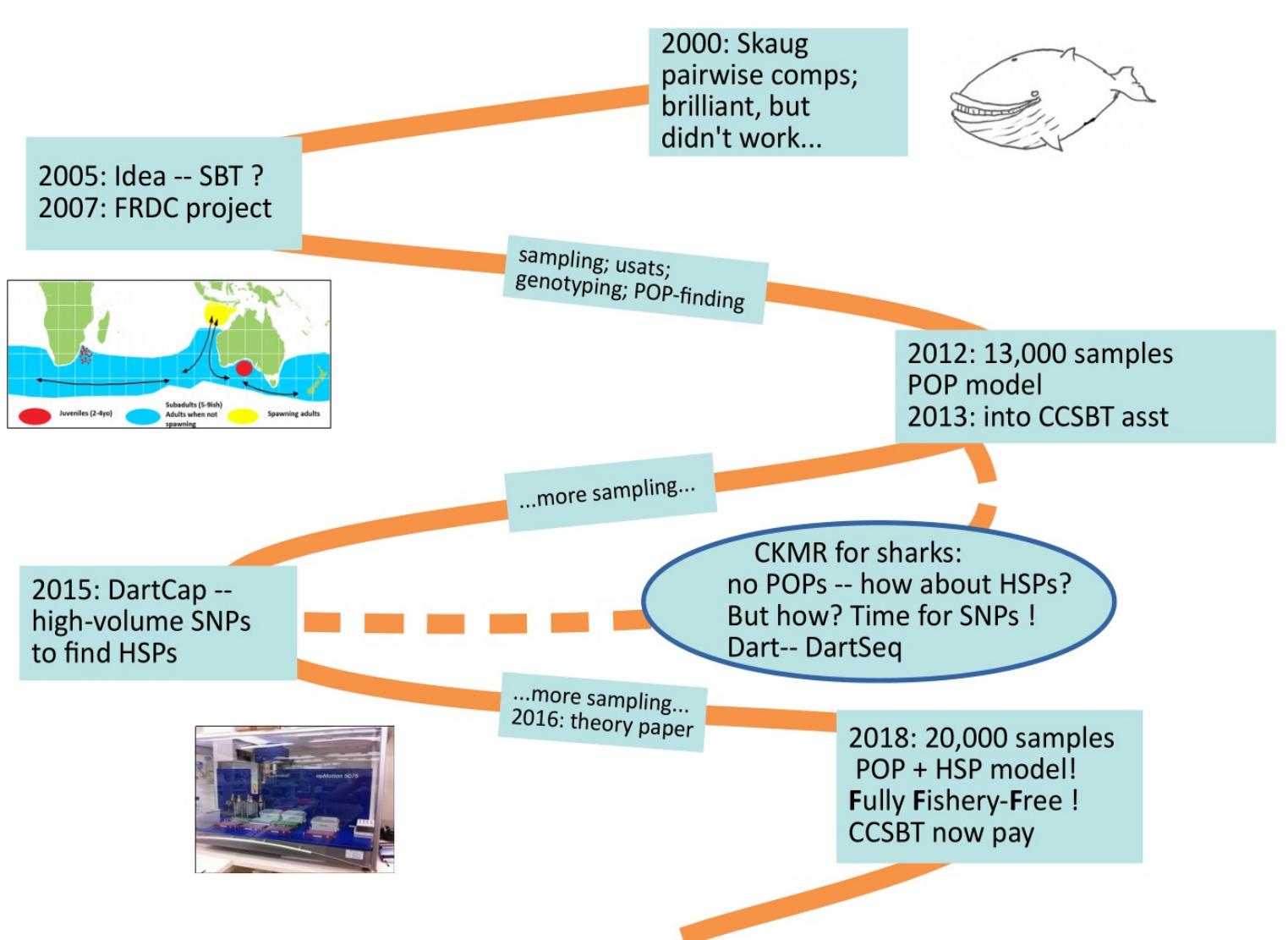
... completed, ongoing, or seriously planned

... NB "starting" is not the same as "succeeding" ...

? 10+ basically done

Species	Source	Geno	Motive	POP	HSP	Model	"Stocks"?	N _{adult}	#Samples
SBTuna	1	Dead	usat	\$	✓	—	Full pop-dyn	—	■,000,000
	2	D	Cap		✓	✓			20,000
School Shark		D	Cap	\$ (choke)	(—)	✓	"	—	■00,000
White Sh.	E	Live+D	Seq	!	—	✓	N, z, ρ	—	■00
	W	L+D	Seq		(—)	✓		—	■000
Grey Nurse Sh.	L+D	Seq	!	✓	✓	Stable-age	—	■000	400
Speartooth Sh.	L	Cap	!?	—	✓		N, z, ρ	(✓)	■000
Different speartooth Sh.	L	Seq	!?	—	✓		N, z, ρ	(✓)	■000

A long and winding road...



A(nother) bit of history

1990s: Norway establishes DNA register of commercial minke whale catch

Schweder: Aha-- CK for abundance!

2001: Skaug publishes one method on Norway minkes

- right idea, wrong example, too early !
- genotypes inadequate, model not general enough
- and Nilssen notes in passing (but needs triads

2005: CSIRO starts SBT CKMR

2011-13: ... starts other T^(threatened) EPS i
(semelparous)

2013: General demography framework (MVB & HJSkaug)

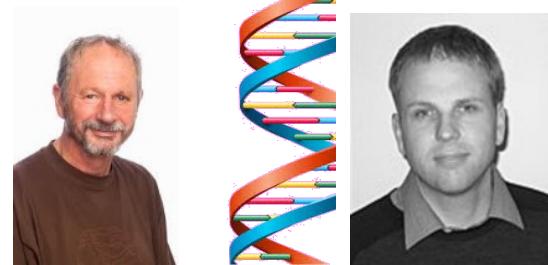
2017-18: framework for design

Other work to date: some stock structure

Individual MR (gene-tagging) now fairly common



BEFORE



AFTER



What you can get from CKMR ...

Biopsies from juves and adults eg from landings plus some size/age info:

absolute abundance of adults

relative **fecundity-at-size** ♀ and ♂

And *if you also know catch-at-age, and have growth curve*

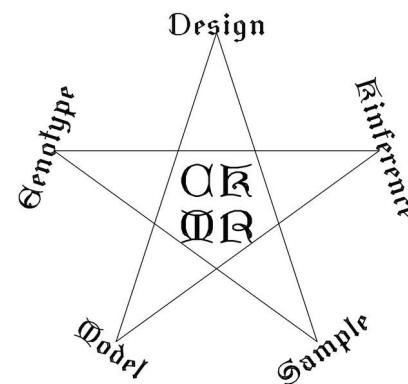
natural mortality averaged across adults

And

connectivity on management timescale (1 generation)

Provided that...

you do everything properly :)



▼

Thus endeth the kinspiel...

... time to buckle up!



ERRO: the key to it all

Expected Relative Reproductive
Output

$\mathbb{P} [\text{Amy is Julian's mum} | \text{stuff about A \& J}]$

$$= \mathbb{E} \left[\frac{\#\text{A's J-like offspring @ } b_J}{\text{Total } \#\text{ J-like offs @ } b_J} | \text{stuff about A} \right]$$

LEXIGRAM GOES HERE...

b: birth-year

y: when yanked from the sea

a: age (usually at *y*)

t: time in general



Factors to always *consider* for ERRO

TIME

SEX of “adult”

AGE of “juve”

AGE and maybe SIZE of “adult”

PLACE

QUIRKS

“adult”: *potential parent*
(seen in POP, unseen in HSP)

“juve”: *potential offspring*
(in POP or in HSP)

“I didn’t measure it” **doesn’t** imply that “it doesn’t matter”
... see Examples!



ERRO for HSPs

Fred is born first; Lucy later

$$\begin{aligned} \mathbb{P} [\text{Fred \& Lucy are MHSP} | \text{stuff about F \& L}] \\ = \mathbb{P} [\text{Lucy has same Mum as Fred} | \text{stuff}] \end{aligned}$$

ERRO for HSPs

Fred is born first; Lucy later

$$\begin{aligned} & \mathbb{P} [\text{Fred \& Lucy are MHSP} \mid \text{stuff about F \& L}] \\ &= \mathbb{P} [\text{Lucy has same Mum as Fred} \mid \text{stuff}] \\ &= \sum_{\text{"Mary" } \in \text{Fred's possible mothers}} \mathbb{P} [\text{Fred's Mum was "Mary"}] \times \dots \\ &\quad \dots \mathbb{P} [\text{"Mary" survived til } b_{\text{Lucy}}] \times \dots \\ &\quad \mathbb{P} [\text{Lucy's Mum was "Mary"} \mid \text{"Mary" alive at } b_{\text{Lucy}}] \end{aligned}$$

ERRO for HSPs

Fred is born first; Lucy later

$$\begin{aligned} & \mathbb{P} [\text{Fred \& Lucy are MHSP} | \text{stuff about F \& L}] \\ &= \mathbb{P} [\text{Lucy has same Mum as Fred} | \text{stuff}] \quad \text{ERRO} \\ &= \sum_{\text{"Mary"} \in \text{Fred's possible mothers}} \mathbb{P} [\text{Fred's Mum was "Mary"}] \times \dots \\ & \quad \dots \mathbb{P} [\text{"Mary" survived til } b_{\text{Lucy}}] \times \dots \quad \text{ERRO} \\ & \quad \mathbb{P} [\text{Lucy's Mum was "Mary"} | \text{"Mary" alive at } b_{\text{Lucy}}] \end{aligned}$$

ie ERRO-weighted average of ERROs

ERRO for HSPs

Fred is born first; Lucy later

$$\begin{aligned} & \mathbb{P} [\text{Fred \& Lucy are MHSP} | \text{stuff about F \& L}] \\ &= \mathbb{P} [\text{Lucy has same Mum as Fred} | \text{stuff}] \quad \text{ERRO} \\ &= \sum_{\text{"Mary"} \in \text{Fred's possible mothers}} \mathbb{P} [\text{Fred's Mum was "Mary"}] \times \dots \\ & \quad \dots \mathbb{P} [\text{"Mary" survived til } b_{\text{Lucy}}] \times \dots \quad \text{ERRO} \\ & \quad \mathbb{P} [\text{Lucy's Mum was "Mary"} | \text{"Mary" alive at } b_{\text{Lucy}}] \end{aligned}$$

ie ERRO-weighted average of ERROs

... breaks down if F & L in *same* cohort

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

$$\begin{aligned} & \mathbb{P} [\text{Amy is Julian's mum} | \text{stuff about A \& J}] \\ = & \mathbb{E} \left[\frac{\#\text{A's J-like offspring } (@b_J)}{\text{Total } \# \text{ J-like offs } (@b_J)} | \text{stuff (ie covariates) of A,J} \right] \end{aligned}$$

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

$$\begin{aligned} & \mathbb{P} [\text{Amy is Julian's mum} | \text{stuff about A \& J}] \\ &= \mathbb{E} \left[\frac{\#\text{A's J-like offspring } (@b_J)}{\text{Total } \# \text{ J-like offs } (@b_J)} | \text{stuff (ie covariates) of A,J} \right] \end{aligned}$$

OK, as long as there's no missing property of "Amys" that affects *both*:

- (i) Amy's ERRO of J-likes among A-likes, *and*
- (ii) Amy's sampling prob relative to other A-likes

after allowing for "Stuff" (which is already included in the model)

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

$$\begin{aligned} & \mathbb{P} [\text{Amy is Julian's mum} | \text{stuff about A \& J}] \\ = & \mathbb{E} \left[\frac{\#\text{A's J-like offspring } (@b_J)}{\text{Total } \# \text{ J-like offs } (@b_J)} | \text{stuff (ie covariates) of A,J} \right] \end{aligned}$$

OK, as long as there's no missing property of "Amys" that affects *both*:

- (i) Amy's ERRO of J-likes among A-likes, *and*
- (ii) Amy's sampling prob relative to other A-likes

after all

- "not measured" does *not* imply "I can ignore it" !
- heritable or non-heritable
- "Stuff" includes: fact of sampledness, and circumstances



"Will it be biased?" HSP Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

$$\begin{aligned} & \mathbb{P} [\text{Fred and Lucy are MHSPs} | \text{F-stuff, L-stuff}] \\ &= \sum_{\text{"Mary"} \in \text{poss F-mums}} (\text{Mary's ERRO of F-likes}) \times \\ & \quad (\text{Mary's future ERRO of L-likes}) \end{aligned}$$

OK, as long as there's no missing property of "Marys" that leads to correlated ERROs of F-likes and L-likes

- see previous



Interpretation of CKMR abund ests

- CKMR abund *is* absolute, ie equiv **number of animals**. But:
 - *which* animals? what “units” ?

This *can* get really subtle; if the next slide confuses you, don’t worry (for now...)

Interpretation of CKMR abund ests

- CKMR abund *is* absolute, ie equiv **number of animals**. But:
 - *which* animals? what “units” ?

Don’t “over-lawyer” this; plus, a properly-set-up model will often give you “what you really want” anyway.

But, *basically*:

Interpretation of CKMR abund ests

- CKMR abund *is* absolute, ie equiv **number of animals**. But:
 - *which* animals? what “units”?

Don’t “over-lawyer” this; plus, a properly-set-up model will often give you “what you really want” anyway.

But, *basically*:

- POPs: *abund of parental generation[*] in units of...*
...“equiv typical sampled adults”
- HSPs: *abund of parental generation in “equiv typical parents”*

[*] IE of adults that gave rise to the “juves” from which you drew your O samples

Fitting a CKMR model

Chassis: a standard-ish Age- and Sex- structured Pop Dyn model

PARAMETERS: "populate" the Pop Dyn Model

$$N_{a+1,t+1,s} = N_{ats} e^{-z}; \quad a \geq 8, \quad t \geq 2002$$

$$\log N_{8t} \sim N(\mu, \sigma^2)$$

$$\log N_{2002,a} \sim N(\mu e^{-z'(a-8)}, \sigma^2)$$

COVARIATES: for each pair of samples (eg "Mary" and "Simon"),

use ERRO to work out POP and HSP probabilities, e.g.

$\Pr[\text{ Mary is Simon's mother } | \text{ Mary's covariates, and Simon's}]$

RESPONSE DATA: the kinship of each pair: POP, HSP, UP (unrelated)

inferred from their genotypes

LOG-LIKELIHOOD: lots of Bernoulli (yes-no) comparisons

Can put in other data too

Fitting a CKMR model

Chassis: a standard-ish Age- and Sex- structured Pop Dyn model

PARAMETERS: "populate" the Pop Dyn Model

$$N_{a+1,t+1,s} = N_{ats} e^{-z}; a \geq 8, t \geq 2002$$

$$\log N_{8t} \sim N(\mu, \sigma^2)$$

$$\log N_{2002,a} \sim N(\mu e^{-z'(a-8)}, \sigma^2)$$

COVARIATES: for each pair of samples (eg "Mary" and "Simon"),

use ERRO to work out POP and HSP probabilities, e.g.

$\Pr[\text{Mary is Simon's mother} | \text{Mary's covariates, and Simon's}]$

RESPONSE DATA: the kinship of each pair: POP, HSP, UP (unrelated)

inferred from their genotypes

$$\text{LGLK}+ = \begin{cases} \log p_{\text{Mary,Simon are xxP}} & \text{M&S really are xxP} \\ \log 1 - p & \text{if not} \end{cases}$$

LOG-LIKELIHOOD: lots of Bernoulli (yes-no) comparisons

Can put in other data too

Two big simplifications

$$\text{LGLK+} = \begin{cases} \log p_{\text{Mary,Simon are xxP}} & \text{M&S really are xxP} \\ \log 1 - p & \text{if not} \end{cases}$$

1. $\Pr[\text{ M&S are Kin }] = \langle\text{something}\rangle / \text{"N"}$

N is big (>100 for sure)

Binomials with small Pr \approx Poisson

2. Lots of “Marys” and “Simons” have same covariates

Poissons with same Pr *add up*

$\Pr[\#\text{Mary-Simon-esque kin-pairs}] \sim \text{Poi}(\#\text{comps_MSlike} * \text{Pr_MSlike})$

So: *group* your comparisons, calculate each Prob *once*

“The Framework”

$$p \sim O(1/N)$$

Binomial with tiny p = Poisson
Poissons add up, to give
Poisson

```
lglk <- function( params) {  
  
  # data prep already done & known about here...  
  
  unpack() ; # split and de-transform parameters  
  
  popdyn() ;  
  
  calc_pure_kinprobs() ; # true kins, ideal covars eg true age  
  calc_obs_kinprobs() ; # distinguishable kins, available covars  
  
  L := 0 ;  
  
  for( K in distinct_kintypes) {  
    L += sum( log_dpoisson( nkins[[K]] ,  
mean=ncomps[[K]] * kinprobs[[K]] ) );  
  } ;  
  
  L += posthoc_pairs() ; # extra info on known kinpairs, eg mtDNA  
  L += other_lglks() ; # age-at-length; etc; priors; CPUEeeeuggggh...  
  
  return( L)  
}
```

