Data Integration Rules - North Carolina

Several questions arose during the data compilation process:

1. If traits data for taxa are available from multiple sources, which source should we use? What if they differ?
2. How do we assign genus-level traits information if only species-level information is available? What if trait states vary among species within the genera?
3. What if traits are co-occurring (more than one trait state is appropriate and is therefore listed)? This was particularly relevant for functional feeding group and habit traits

Integration rules were developed to maintain consistency when addressing these issues. For most of the traits, the Poff et al. 2006 Traits Matrix was given top priority. If the Traits Matrix lacked information for certain taxa, the USGS traits database (Vieira et al. 2006) received next highest priority, followed by the U.S. EPA's 1970s publications. Weighted average and maximum likelihood calculations received top priority for the temperature preference and tolerance trait assignments. All OTUs in the state biomonitoring databases, including rare taxa, were included North Carolina traits table. This is because the database is meant to be a 'living' document; the intent is that people using the database can fill in or update information as it becomes available. People using the database are encouraged to check the trait information and customize it as necessary so that the information is more accurate for taxa occurring in their region (in particular functional feeding group (FFG) and habit, for which only primary trait state assignments were made).

The trait information that was entered into the North Carolina traits table came from a number of different sources. Sometimes the sources had slight differences in how traits were categorized and in some of the thresholds that were used when assigning trait states. Another issue was that trait information for certain taxa was available from several different sources, so a decision had to be made about which source to use (sources were generally in agreement, but sometimes slight differences existed). Because of these issues, decisions had to be made during the entry process. One involved interpreting literature in order to get the trait state information into a standardized and usable format for analyses. The other involved deciding which source to use. 'Rules' were developed for the following trait state entries: voltinism, development, life span, dispersal, armoring, size, rheophily, functional feeding group, habit, tolerance values and thermal preference and tolerance. They are summarized in **Tables 1** through **11**.

Table 1. Integration rules that were used when assigning voltinism trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Voltinism	Poff et al. 2006	semivoltine	semivoltine
		univoltine	univoltine
		bi- or multivoltine	bi- or multivoltine
Voltinism	Vieira et al. 2006	< 1 Generation per year	semivoltine
		1 Generation per year	univoltine
		> 1 Generation per year	bi- or multivoltine

Rules:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entries

Many of the Vieira entries went to species-level. If trait states varied among species within a genus:
- a. the trait state that was most frequently recorded was used (= majority rules)
- b. if different trait states occurred with the same frequency, the Volt_Comments field was referenced. If it mentioned that one state was more typical than another, the more typical state was used.
- c. if Volt_ comments was not helpful, the trait state with the higher number of generations was chosen. For example, if there was one 'univoltine' entry and one 'semivoltine' entry, the 'univoltine' entry was chosen.

Table 2. Integration rules that were used when assigning development trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Development	Poff et al. 2006	Fast seasonal	Fast
		Slow seasonal	Slow
		Non-seasonal	Non
Dev_Speed	Vieira et al. 2006	Fast seasonal	Fast
		Slow seasonal	Slow
		Non-seasonal	Non

Rules:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entries

Table 3. Integration rules that were used when assigning life span trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Adult Life Span	Poff et al. 2006	very short	very short
		short	short
		long	long
Adult_lifespan	Vieira et al. 2006	hours	very short
		days	very short
		weeks	short
		months	long

Rules:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entries (reference Adult_lifespan_comments if necessary)

Table 4. Integration rules that were used when assigning dispersal trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Female dispersal	Poff et al. 2006	low (< 1 km flight before laying eggs)	low
		high (> 1 km flight before laying eggs)	high
Adult dispersal	Vieira et al. 2006	1 km or less	low
		10 km or less	high
		10 m or less	NA
		100 km or less	high

Rules:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entries

Notes:
in the Poff et al. 2006 table, it specifies 'female dispersal.'
in the Vieira et al. 2006 traits database, it specifies 'Adult dispersal.'
It was assumed that the information was compatible between sources.
In Vieira et al. 2006, there is an entry '10 m or less.'
It appears that this was a typo (it likely should have been '10 km or less').
Therefore this category was excluded.

Table 5. Integration rules that were used when assigning armoring trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Armoring	Poff et al. 2006	none (soft-bodied forms)	none
		poor (heavily sclerotized)	poor
		good (e.g., some cased caddisflies)	good
Armor	Vieira et al. 2006	Soft	none
		All sclerotized	poor
		Partly sclerotized	poor
		Hard shelled	good

Rules: 1. use the Poff et al. 2006 entries (for genus-level matches)

2. use the Vieira et al. 2006 entries

Notes: In the Poff et al. 2006 table, it does not mention 'partly sclerotized.'

In the Vieira et al. 2006 table, 'partly sclerotized' and 'all sclerotized' were assigned to the 'poor' category.

Table 6. Integration rules that were used when assigning size (at maturity) trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Size at maturity	Poff et al. 2006	Large (length > 16 mm)	large
		Medium (length 9-16 mm)	medium
		Small (length < 9 mm)	small
Max_Body_Size	Vieira et al. 2006	Large (length > 16 mm)	large
		Medium (length 9-16 mm)	medium
		Small (length < 9 mm)	small

Rules: 1. use the Poff et al. 2006 entries (for genus-level matches)

2. use the Vieira et al. 2006 entries

If more than one trait state was assigned (i.e. there was variation among species within a genus):

a. the category that was most frequently recorded was used (majority rules)

b. if different categories were recorded the same number of times, the 'medium' entry was used (i.e. if there was one 'small' entry and one 'medium' entry, the medium entry was used)

Table 7. Integration rules that were used when assigning rheophily trait states to taxa.

Trait	Source	Original trait states	Assigned trait states
Rheophily	Poff et al. 2006	Depositional only	depo
		Depositional and erosional	depo_eros
		Erosional	eros
Rheophily	Vieira et al. 2006	Current_quiet	depo
		Current_slow	depo
		Current_fast_lam	eros
		Current_fast_turb	eros
		More than one:	if both quiet and slow, depo
		quiet & slow	depo
		quiet and/or slow and fast (either lam or turb)	depo_eros
Flow_pref	EPA 1970s	Standing	depo
		Slight	depo
		Standing-slight	depo
		Standing and flowing	depo_eros
		Moderate	eros
		Moderate-fast	eros
		Fast	eros
		More than one:	
		some combination of standing and/or slight and moderate and/or fast	depo_eros

Rules:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entries
3. use the EPA1970s entries

If more than one trait state was assigned (i.e. there was variation among species within a genus):
the category that was most frequently recorded was used (majority rules)

Table 8. Integration rules that were used when assigning (primary) functional feeding group trait states to taxa.

Integration Rules for FFG:
Only one FFG category was assigned to each taxa. The following rules were used:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entry (Feed_mode_prim)
3. use the WSA entry from the Benthics_Master_Taxa table
4. use the RBP2 1999 entry from the Benthics_Master_Taxa table
5. use the USEPA 1990 Draft entry from the Benthics_Master_Taxa table
If more than one category was assigned within a genus, the one that occurred most frequently was entered (=majority rules)
If different states were recorded the same number of times, the next source was used as a 'tie-breaker' (i.e. if Vieira et al. 2006 had two species listed as CN and two as SP, and the WSA entry was SP, I used SP)
If unable to resolve based on these sources, one was randomly selected.

Table 9. Integration rules that were used when assigning (primary) habit trait states to taxa.

Integration Rules for Habit:
Only one habit category was assigned to each taxa. The following rules were used:
1. use the Poff et al. 2006 entries (for genus-level matches)
2. use the Vieira et al. 2006 entry (Habit_prim)
3. use the WSA entry from the Benthics_Master_Taxa table
4. use the RBP2 1999 entry from the Benthics_Master_Taxa table
5. use the USEPA 1990 Draft entry from the Benthics_Master_Taxa table
If more than one category was assigned within a genus, the one that occurred most frequently was entered (=majority rules)
If different states were recorded the same number of times, the next source was used as a 'tie-breaker' (i.e. if Vieira et al. 2006 had two species listed as CN and two as SP, and the WSA entry was SP, I used SP)
If unable to resolve based on these sources, one was randomly selected.

Table 10. Integration rules that were used when assigning tolerance values to taxa.

Integration Rules for Tolerance:
Only one tolerance value was assigned to each taxa. The following rules were used: 1. use the WSA entry 2. use the RBP2 1999 entry 3. use the USEPA 1990 Draft entry If there were more than two values from a source, the median value was used. If there were two entries, the higher value was used (i.e. if assigned values were 2 & 3, the 3 was used) NOTE: if state-specific tolerance values were provided, those were also incorporated into the traits table.

Table 11. Integration rules that were used when assigning thermal preference and tolerance values to taxa.

Traits	Source	Original trait states	Assigned trait states
Thermal preference	Poff et al. 2006	cold_cool	Rank_opt = 3, Rank_tol=3
		cool_warm	Rank_opt = 4, Rank_tol=5
		warm	Rank_opt = 5, Rank_tol=3
Thermal_pref	Vieira et al. 2006	Cold stenothermal (<5 C)	Rank_opt = 3, Rank_tol=3
		Cold-cool eurythermal (0-15 C)	Rank_opt = 3, Rank_tol=4
		Hot euthermal (>30 C)	Rank_opt = 5, Rank_tol=3
		No strong preference	Rank_opt = 4, Rank_tol=5
		Warm eurythermal (15-30 C)	Rank_opt = 5, Rank_tol=4
		More than one: combination of colder and warmer categories	Rank_opt = 4, Rank_tol=5
Thermal preference	EPA 1970s	Eurythermal (≥15 C)	Rank_opt = 5, Rank_tol=4
		Euthermal (>30 C)	Rank_opt = 5, Rank_tol=3
		Mesothermal (15-30 C)	Rank_opt=5, Rank_tol=4
		Metathermal (5-15 C)	Rank_opt=3, Rank_tol=3
		Oligothermal (<15 C)	Rank_opt = 3, Rank_tol=4
		Stenothermal (≤ 5 C)	Rank_opt = 3, Rank_tol=3
Temp_Opt_Rank		Wide range - no apparent preference	Rank_opt = 4, Rank_tol=5
Temp_Tol_Rank		More than one: combination of colder and warmer categories	Rank_opt = 4, Rank_tol=5

Rules:

1. use the values generated by Lei Zheng (or from other databases, like Brandt and Yuan)
1. use the Poff et al. 2006 entries (for genus-level matches)
3. use the Vieira et al. 2006 entries
4. use the EPA1970s entries

If more than one trait state was assigned (i.e. there was variation among species within a genus):
the category that was most frequently recorded was used (majority rules)