Table 1

Watershed characteristics

Transformation and pathways

Sources and sinks

Proxy indicators

Surface vs sediment

(1) Qu

Do core/satellite molecule different latitudes

What watershed factors relate to the molecules?

Are there always 'core' molecules?

Do core/satellite molecules vary along the fluvial network?

Is there a tendency towards clustering of molecules? That is do the collection of molecules becomes more alike along a flow path/ through time.

re satellite nolecules iologically ecalcitrant?

What is the relationship between connectivity (in the putative transformation network) and thermodynamic properties (delG, lambda) in core and satellite metabolites?

Do the molecules indicate transformation occurred?

How does the niche breadth of molecules vary?

Are the molecules terrestrially

Would the core molecules remain even with disturbance?

Do the molecules indicate presence/absence of other biota? Like endangered species?

Biota in river networks display same diversity as the diversity of the molecules

How do molecules differ across sediment and water column samples?

(2) Hypotheses

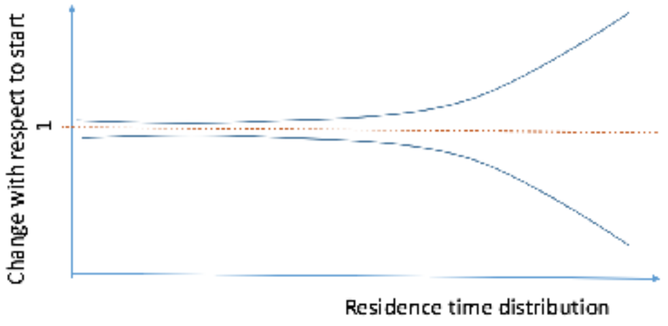
What is the relative diversity of molecules between core and satellite molecules?

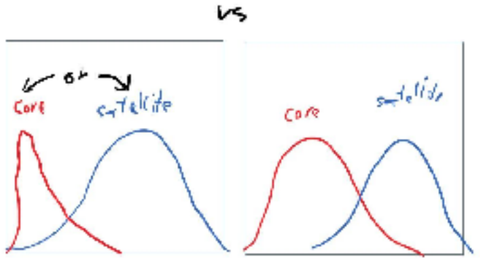
Can the relative diversity of olecules be lated to residence ne distribution?

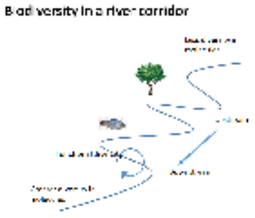
Are there differences in lability among core and satellite molecules? Does this change along the network?

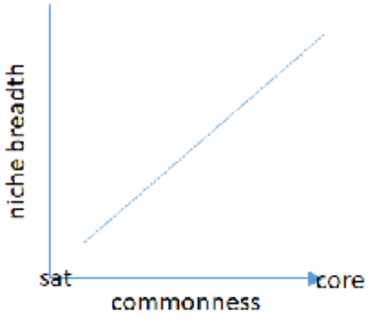
There is a tendency olecule

More widespread molecules are "generalists" with wider niches.









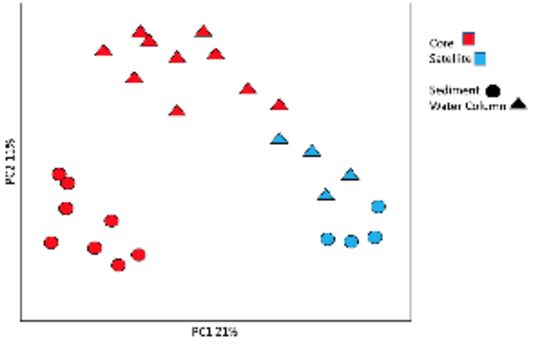


Table 2

Jay Z, Nathan W, Amy B, Jess, Wai-Yin

Scale-dependence

Q: Ancillary/meta data?

Biogeo transform

SW/sed variation

(1) Questions/concepts

What is the scale-dependence of core metabolites? Are some molecules only "core" at the local-scale? Are others regional or global core molecules?

Is it scale dependent with variation depending on more local landscape/river network mosaic composition (e.g., density of wetlands/lakes/urban/ gaining-losing river)?

How important is recent history (antecedent conditions) to controlling interpretation of Core vs Satellite? Does this obscure spatial patterns?

Are core molecules precursor material or the broken down remains / simpler molecules that result from in-stream transformations?

Hyp: Core molecules will have higher lambda because they are less favorable for microbial production

Because there are transformations in the biochemical network, are there also core "links" that we can detect in the data, common among sites?

How do core and satellite molecules differ in general across space? Defining characteristics? Similarities found across space? Aquatic vs terrestrial sources?

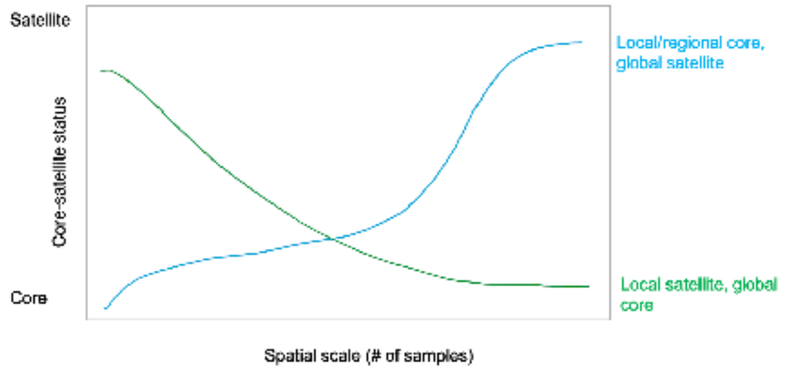
How does the GFE differ between core and satellite in sediment vs. surface water? H: Higher GFE in sediment, but likely varies by ecoregion.

(2

How does the GFE of molecules in the core vs. satellite components shift along the stream to river continuum? (do we have any nested watersheds to look at this?)

Sites underlain by permafrost will have more homogeneous FTIR characteristics and more core and less satellite compared to temperate sites. Maybe similar sorting in other extreme regions?

(3) predictions/figures



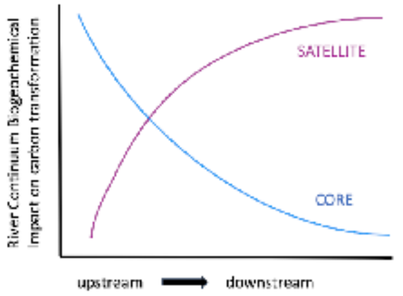


Figure C. (A) hypothesized impacts of space (across a watershed) on biogeochemical transformation of carbon.

Q: How does the GFE differ between core and satellite molecular components in surface water vs. sediment?
H1: Low GFE in sediment, but controlled by ecoregion.
H2: Difference between core vs. satellite is greater in sediment than in water.
H3: Ratio of Water:Sediment GFE (within core or sat subset) will be

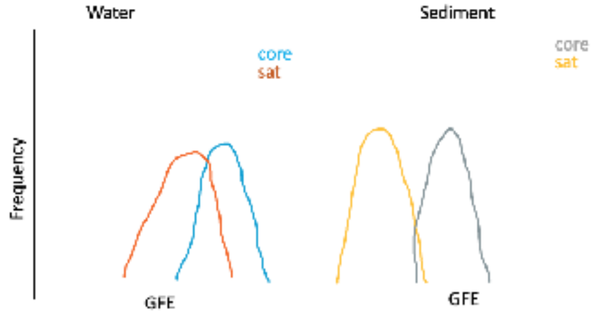


Table 3

Landscape and larger scale

Molecular properties and reactivity

How do changes in landscapes catchment inputs (e.g., from varying surface features, hydrology, geology) affect the diversity of satellite molecular compounds in surface water moving from headwaters to river outlet?

Are there major differences in core DOM compounds/classes across geographic locations and river sizes?

How does size of core molecules in sediment and surface water compare to each other?

How does reactivity of compounds change (ie NOSC, AI) with CORE vs SATELLITE (maybe also SURFACE vs POREWATER)?

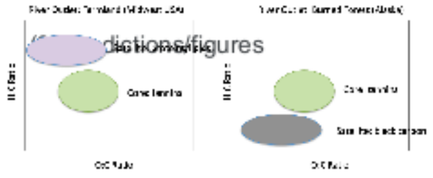
--- effects >

Watersheds that contain gradients/shifts in landscape characteristics diversify satellite compounds and classes in surface water compared to those with more uniform landscape features.

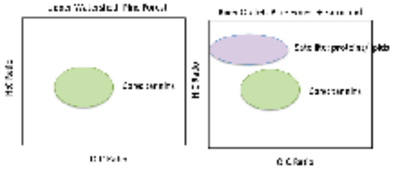
We expect locations with different broader landscape types (e.g., tundra vs. deciduous forest) contain different core compounds in surface waters.

size of core molecules are smaller in sediment than surface water

NOSC: CORE has less labile (conserved) compounds



dictions/figures



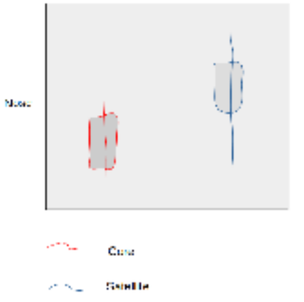


Table 5

(1) Questions/concepts

Is there are chemical structure similarity among core vs satellite?

universal vs. region- or river type specific vs. site specific (taking core vs. satellite further)

"niche breadth" of metabolites

environmental drivers of metabolites and/or biochemical transformation networks

What are microbial-vegetation-hydrology relationships we can discover with WHONDRS and what does it tell us about what drives site similarity?

Do the rules of core vs sat vary by dominant vegetation?

How do thermodynamic properties and biochemical connectivity characterize core and satellite molecules?

What does it mean to be a "core" or "satellite" molecule?

(2) Hypotheses

Conversion of core molecules will be more thermodynamically favorable

Core molecules will have higher connectivity than satellite molecules

more recalcitrant compounds found further downstream

core molecules are leftover = recalcitrant and form a background, while satellite molecules are dynamic and more labile.

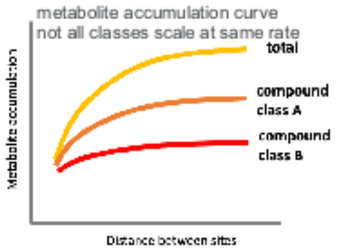
(3) predictions/figures

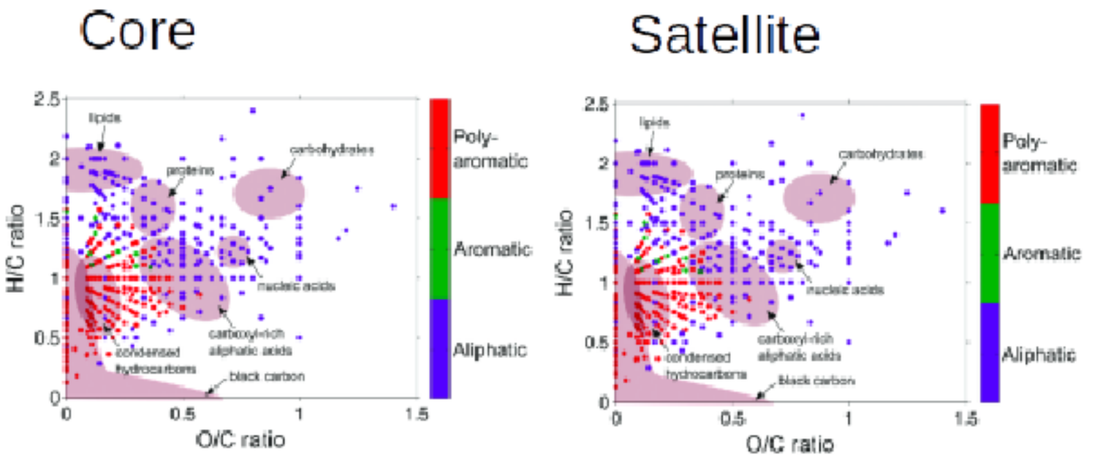
"niche breadth" of metabolites

what drives the differences between core and satellite molecules (e.g., land cover, androgenic activities, watershed characteristics)?

universal vs. region- or river type-specific vs. site-specific (furthering core vs. satellite categorization)

characterize core vs satellite by landscape/vegetation/ region

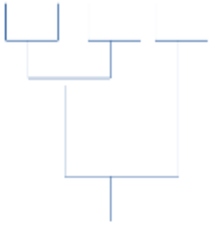




What does it mean to be a "core" or "satellite" molecule?

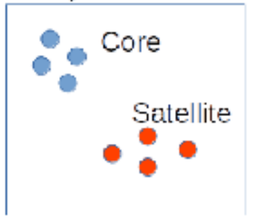
is there a chemical structure similarity among core vs satellite?

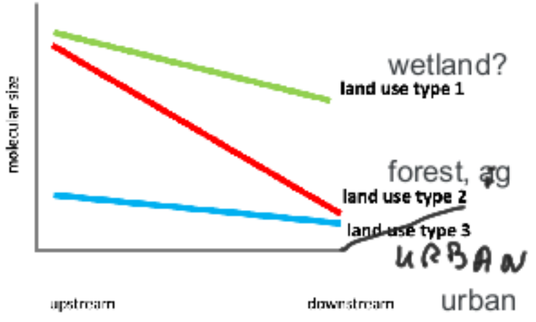
core molecules are leftover molecules (recalcitrant), satellite are more labile and dynamic



Phylogeneti c tree with chemical structure

Unifrac based multivariate analysis





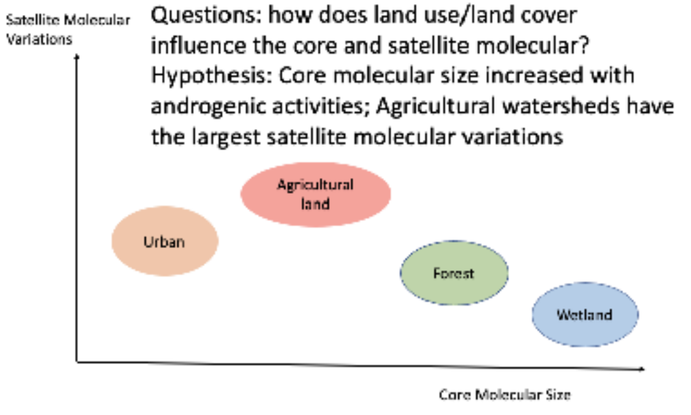


Table 7

Baseflow-core within a core

Is there an environmental/locational linkage between # of core vs # satellite compounds?

Can everyone see this note?

(1) Questions/conce

Questions about stoichiometry and bioavailablity in the Core vs. Satellite Compounds

Questions about the number of heterogeneous molecular formuale increases for satellite data vs core

+1

Stoichiometry variability

Linkages to site metadata to determine core, satellite data etc - does the core change based on underlying characteristics?

How does the distribution of these N and S molecules vary over space?

How do nutrient ratios vary across low to high carbon environments?

add metadata to sampling locations

(2) Hypothe

H: The number of heterogeneous molecular formulae is greater for satellite data than core data because heterogeneity is driven by ecosystem diversity.

please edit: samples taken at base flow --> prob?

Timing of sampling aspect REALLY important!

(3) predictions/figures

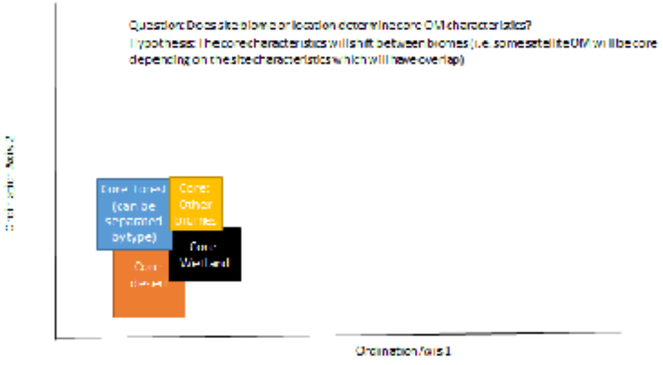


Table 8

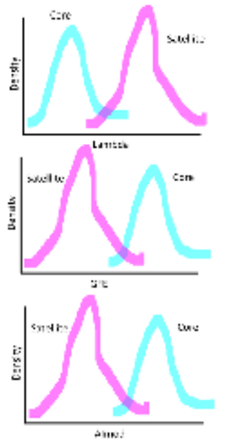
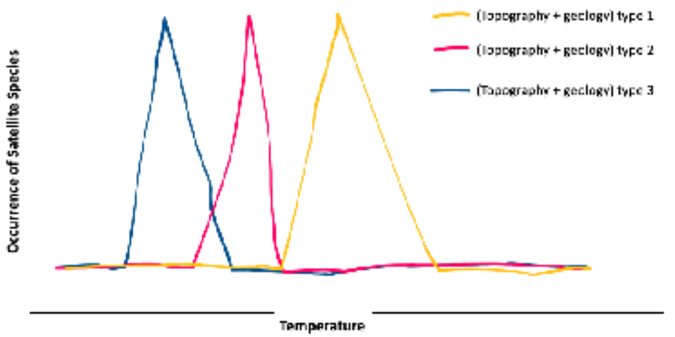
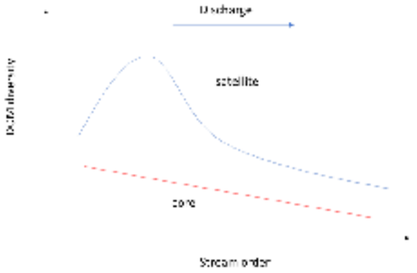
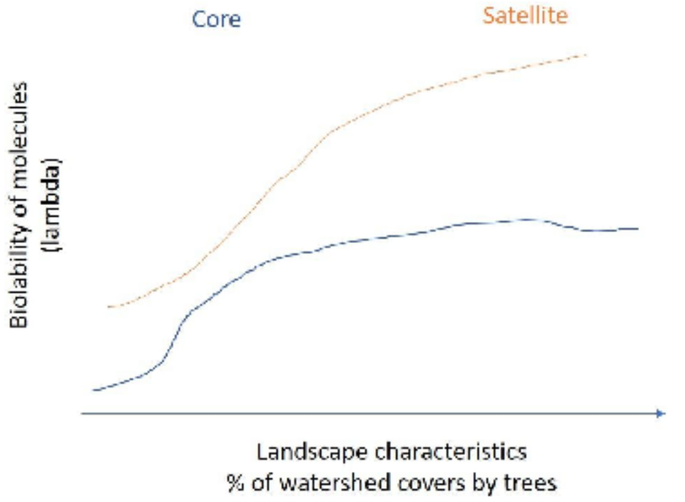
	Landscape + climate drivers	DOM properties	Inherent chemical components	Microbial drivers
(1) Questions/concepts	What watershed/water characteristics are driving the occurrence of satellite species in particular?	How do molecular mass and other physical and functional properties of differ between core and satellite molecules?		biosphere. Rare biosphere taxa, with the right resource pulse, can experience a boom in growth, but can also experience a rapid bust due to viral lysis. linking this to compounds, then perhaps satellite compounds are also those that have rapid Like Erika's, but specifically about GFE and lambda: Does the GFE and lambda (according to James was a measure of efficiency of microbial use for biomass (?)) differ between core and satelite molecules? Hypothesis almost depends on turnover and production, and what its ubiquity means (is it persistent and that's why it's ubiquitous) or is it ubiquitous because it has a wide range of sources and production of these molecules is rapid
(2) Hypotheses	How does the core/satellite concept relate to the existing ecological conceptual frameworks	The temperature might be the direct driver, the impact of which is likely modulated by the topography and geology.		
(3) predictions/figures	Relative importance of landscape features versus in-stream processes and contributions if microbial community			

Table 9

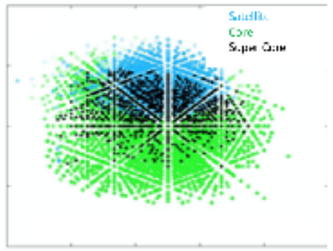
All samples at base flow.

(1

Source of core vs satellite - terrestrial vs. plankton/microbial source

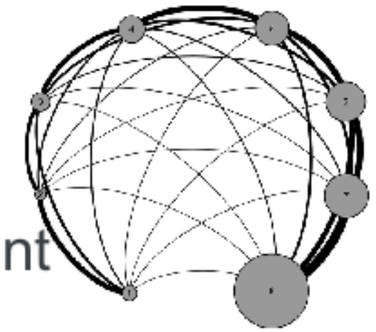
How does the balance of autochthonous production vs allochthonous inputs at base flow influence molecular composition of DOM pool?

Rivers with higher light exposure (i.e. less shading) and light penetration (i.e. shallow or non-turbid) wil have a greater proportion of algal derived DOM



(3) predictions/figures

From Raymond 2016: Pulse-Shunt Concept



How to group watersheds from chemical composition, cluster analysis, ordination?

Are Satellite species more transient than Core?

How does the thermodynamic influence vary between core and satellite molecules?

How does groundwater source affect water chemistry?

Balance of groundwater vs surface water contributions and history

Is there a clear separation between Core and Satellite? What does it mean if there is, or if there isn't?

How does the residence time of a river system influence DOM transformations?

The thermodynamics influence increases as a saturating function of C limitation. Given that satellite molecules are diverse in H_1C and O_1C ratio, the thermodynamic influence is probably lower than the core molecules.

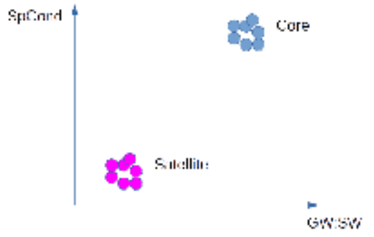
What environmental variables are correlated with chemical variation

How does proportion of core/satellite change with stream order?

How does core vs satellite composition change during high flow events?

Longer residence time allows greater extent of molecular transformations. This will vary by stream order and elevation steepness

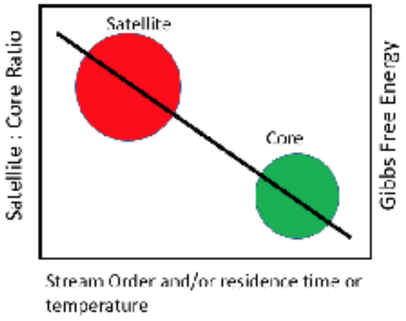
Satellite molecules induce greater microbial stimulation.

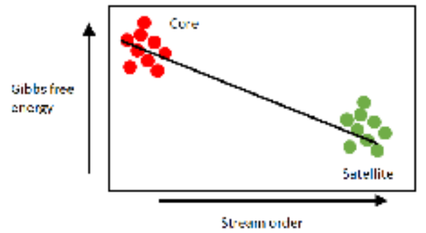


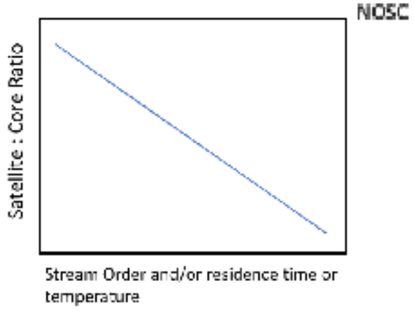
Size of satellite pool decreases with increasing stream order.

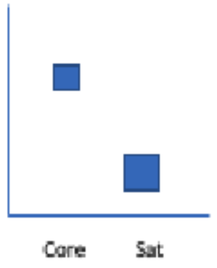
Core molecules are more oxidi

Highly reactive fraction may be "satellite" and present mostly in rivers with short









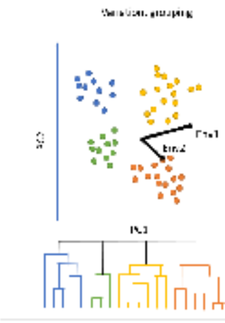


Table 11

Q1: Geospatial

(1) Question

Are there any scaling patterns with respect to stream order (for both core and satellite information)?

To what extent do core and satellite differ by region?

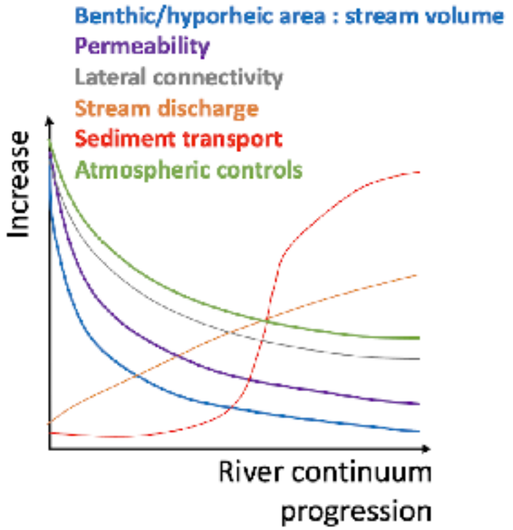
Do patterns between core and satellite information vary with respect to (any information that can be collected about) atmospheric conditions?

(2

Atmospheric controls on sample composition are more important in headwaters than in low-land rivers.

Using machine learning to develop scaling patterns with respect to stream-order and atmospheric conditions (or using metadata from 97 locations)

(3) predictions/figures



Increase/QoI = ML-model(River continuum aggression)

Q2: Microbial Processes

compare oxidation states? compare/relate to catchment characteristics?

Is there a relationship between the core and satellite metabolites and microbial functional groups (groups using different electron acceptors) ?

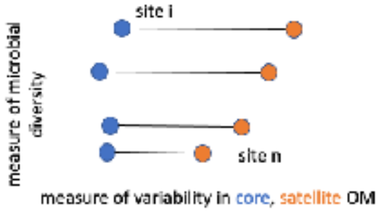
patterns in the difference between sfc and subsfc characteristics across sites. Compare with riverbed characteristics

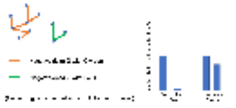
H: Low order streams show a higher oxidation state of the core organic matter

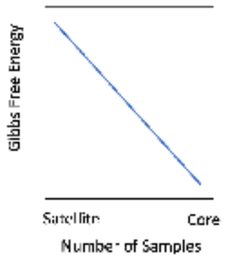
H: a more diverse microbial community coincides with a more variable satellite but not core OM

Connectivity between nodes will be related to thermodynamic favourabillity of reactions when coupled to different electron acceptors.

H1: Molecules with low Gibbs Free Energy are part of the Core because it gives taxa that can use these molecules a competitive edge. H2: Low GFE molecules are in high demand so uncommon.



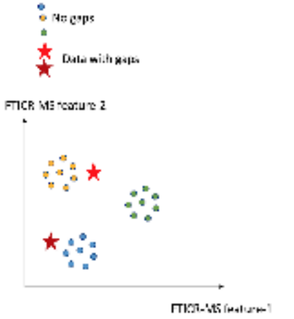




Q3: Methods

How to fill in gaps in the FTICR-MS data?

Using machine learning to better understand the molecular mass difference across core and satellite spectrum, discover patterns, and fill in gaps



Unsupervised ML (e.g., custom k-means clustering, NMFk, PCA, etc)

Table 15

Are there differences between surface and sediment molecules? are these differences greater for core or satellite molecules?

can we predict molecules using widely available remotely sensed data?

from different subsurface depths. Stream water at low flow mostly comes from deeper groundwater; stream flow at high flow comes from mostly soil water. The DOM at low and high flows therefore have very

is it easier to predict surface or sediment molecules using remotely sensed data?

(1) Questions/concepts

Are there difference molecules in the sediment based on different redox conditions? How do they differ across the redox regime?

What are the differences between DOM composition under wet and dry climate, and under natural and human impacted places (ag, urban, deforestation)?

(2) Hypotheses

Is it possible to represent the bulk or average properties of DOM in reactive transport models and still capture the key DOM transformation?

(3) predictions/figures

Surface water and sediments, only space, not time varying

Table 10

Questions/concepts

Can you predict soil composition from surface composition at the same site?

are there samples collected at the same place over time ? could they be used to link to microbial processes ?

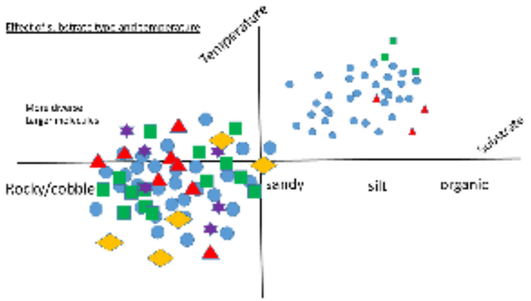
Do the patterns /compositions change along latitude /longitude? Is different light intensity (photodegradation) reflected across the sites ?

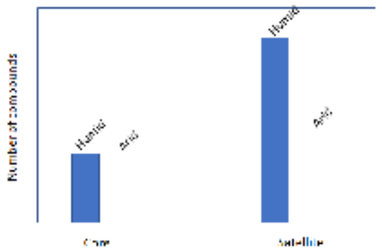
Are their bioindicators (specific molecules) in the data that reflect landuse ?

Do we have information/data about the type of substrate sampled for sediments? rocky, sandy, silty?

Stream waters at low flow mostly comes from deeper groundwater; stream flow at high flow comes from mostly soil water. The DOM at low and high flows therefore have very different composition and diversity.

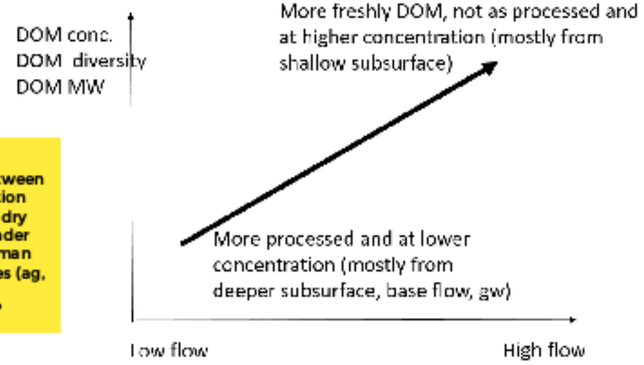
C-N ratios will tell you how processed the organic matter is. Also 13C data.





How will organisms on the sediment surface (e.g. algae) confound this analysis? Is there a way to group out those organisms from the sediment analysis?

What are the differences between DOM composition under wet and dry climate, and under natural and human impacted places (ag, urban, deforestation)?



(2) Hypotheses

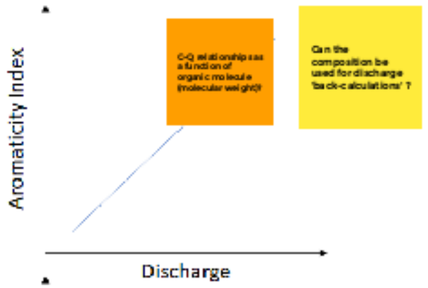
Are their bioindicators (specific molecules) in the data that reflect landuse ?

MW of sediment-derived NOM will be higher than surface water

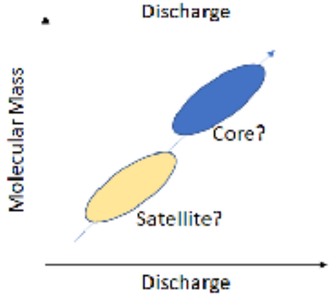
Core compounds will reflect large-scale watershed properties; satellite will be controlled by local land use.

Average lambda of core compounds will be lower than satellite compounds

Humid environments will have higher diversity (more satellite compounds)

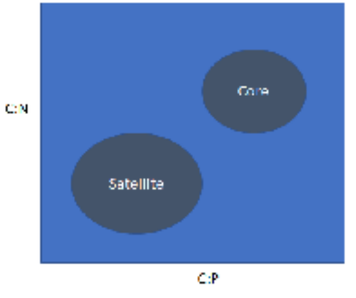


Higher molecular weight compounds will be more prevalent at higher discharges



Question: How do elements other than CHO (P,N,S) differ between core and satellite metabolomes?

Hypothesis: The amount of N reflects the degree of processing and will be depleted in core metabolomes. P is needed for biomass building and will also be depleted in the core metabolome.



(3) predictions/figures