Webb et al Biogeosciences

This is a very interesting paper that addresses biogeochemistry among five different substrate types in the Caribbean using closed in situ tent incubations. There is a wealth of data here that would be great to see published in Biogeosciences. However, I have several key concerns in the methods that should be addressed before this work is considered, mostly relating to the low units of replication, and consequent analyses, narrative and overall extrapolation of reef functioning thereafter.

It seems that the inverse model outputs employed distract from more targeted analyses and presentation of, what could be, quite compelling results on biogeochemical fluxes in the tents among substrate types. This modelling approach may have been adopted owing to the low sample size in the study; n=3 daytime and n=2 nighttime per substrate type. If so, the authors should express this clearly. While I cannot speak for the models themselves, as this is far from my expertise, I am concerned that the models were established based on just 2-3 replicates per treatment. In fact, calculations of 95% CIs (e.g. Fig 6) were also inferred from model outputs owing to low sample size. It is likely that the low level of replication within each substrate type (n=2-3) skews the overall result towards “no significant differences between any communities” (Ln 243), when this is not actually the case. This appears to be a key take-home message (e.g. abstract Ln 28; “no significant difference between processes on any assemblage”) that may not be true, but there was not enough power to show otherwise. This forms a narrative and conclusion that may dangerously misrepresent the data, which poses significant risk when extrapolating findings to ecosystem functioning and functional redundancy.

The inverse modelling seems to overlook specific results among treatments. Table S2 presents model output data, but this does not compare treatments/factors (day vs. night, substrate types). Such comparative analyses are found in Table S3, which is the PERMANOVA result that shows no significant differences, except for day vs. night. It is unclear what data (factors or response variables) were even used for the PERMANOVA, which I feel may not be the correct approach to analysing these data.

I would argue that if more targeted analyses (e.g. linear models, ANOVA) were conducted, say for NCP or NCC between day-night and substrate types, more informative results would emerge. For example, in Fig 6, NCP is much lower at night in coral, BES and BCM than in sand and TMA. Conversely, NCC is greater in the day for coral, BES and BCM than in sand, and lowest at night for BES and BCM. I am just eyeballing data here, but this does not look like “no significant differences between any communities” (Ln 243). I am not sold by the PERMANOVA, as it seems there should be some detectable differences between substrate types, and that greater detail and resolution could uncover this.

Also, it seems that the dominant substrate type was used to categorise factors, but this may impact (and limit) the results due to low sample size. One alternative to this issue could be to analyse all tents using continuous data for substrate type. E.g. could data be analysed at the level of “percent cover of sand”, “percent cover of coral”, “percent cover of cyanobacteria”, and so on… rather than treating them as fixed categorical factors? This would increase sample size, and possibly tease out interesting results e.g. thresholds of cover for positive or negative results regarding ecosystem processes and functions. Otherwise, more data / replicates may be needed. I fear that in using the current approach, conclusions are made on ecosystem functioning and redundancy that extend beyond the scope of the data.

Abstract
Ln 23-33: I suggest changing to past tense here, e.g. “Estimated processes **were** low” and “No real gain in primary habitat **was** recorded”…. And so on.

Ln 28: Suggest removing reference to the analysis here “A multivariate pairwise analysis revealed that there is no significant difference…” to be more succinct, e.g. “We found no significant difference…”

Introduction

Ln 36: “habita**n**t” should be “habitat”

The authors should consider other relevant literature on coral reef ecosystem functioning and functional redundancy, e.g.

- Range of work by Bellwood, e.g.:
    - Bellwood, D. R., Hoey, A. S., & Choat, J. H. (2003). Limited functional redundancy in high diversity systems: resilience and ecosystem function on coral reefs. *Ecology letters*, *6*(4), 281-285
    - Bellwood, D. R., Streit, R. P., Brandl, S. J., & Tebbett, S. B. (2019). The meaning of the term ‘function’ in ecology: a coral reef perspective. *Functional Ecology*, *33*(6), 948-961
- Wolfe, K., Anthony, K., Babcock, R. C., Bay, L., Bourne, D. G., Burrows, D., ... & Mumby, P. J. (2020). Priority species to support the functional integrity of coral reefs. *Oceanography and Marine Biology*.

Ln 70: Additional work could be considered including in situ and lab experiments, e.g.

- Albright, R., Caldeira, L., Hosfelt, J., Kwiatkowski, L., Maclaren, J. K., Mason, B. M., ... & Caldeira, K. (2016). Reversal of ocean acidification enhances net coral reef calcification. *Nature*, *531*(7594), 362-365
- Albright, R., Takeshita, Y., Koweek, D. A., Ninokawa, A., Wolfe, K., Rivlin, T., ... & Caldeira, K. (2018). Carbon dioxide addition to coral reef waters suppresses net community calcification. *Nature*, *555*(7697), 516-519
- Dove, S. G., Kline, D. I., Pantos, O., Angly, F. E., Tyson, G. W., & Hoegh-Guldberg, O. (2013). Future reef decalcification under a business-as-usual CO2 emission scenario. *Proceedings of the National Academy of Sciences*, *110*(38), 15342-15347
- Dove, S. G., Brown, K. T., Van Den Heuvel, A., Chai, A., & Hoegh-Guldberg, O. (2020). Ocean warming and acidification uncouple calcification from calcifier biomass which accelerates coral reef decline. *Communications Earth & Environment*, *1*(1), 1-9
- Brown, K. T., Bender-Champ, D., Achlatis, M., van Der Zande, R. M., Kubicek, A., Martin, S. B., ... & Hoegh-Guldberg, O. (2021). Habitat-specific biogenic production and erosion influences net framework and sediment coral reef carbonate budgets. *Limnology and Oceanography*, *66*(2), 349-365

Methods

Ln 92, 93, 95, 144, 150, 184, etc… Suggest making methods section past tense e.g. change “is” to “was”, “are” to “were”, etc.

Ln 102: More information should be provided on the number of replicate tents used per substrate type;

- There were five substrate treatments, each with three replicates (?) (Fig 2). Were all of these deployed at the same time or was a single custom tent reused for all?
- How long were tents left over the substrate before beginning the experiments / incubations?
- Was the substrate left to stabilise in cases where substrate was moved into the tent to artificially construct the benthic community?
- Daytime incubations were done in triplicates (Ln 102) and night incubations in duplicates (Ln 102), but is this n=3 tents per substrate incubated 3 (day; n=9) and 2 (night; n=6) times **or** just one tent done n=3 (day) and n=2 (night) times?

- If the former, was tent number incorporated as a factor to account for psuedoreplication of tent/substrate type? And were there differences detected in seawater parameters across incubations (i.e. repeated measures?)
- If the latter, units of replication per treatment and timepoint are quite low.

Ln 103: Incubations went for 4 hrs each, but did they all start around the same time day and night? It seems in Fig 5 that O2 does not follow the trend line selected, but instead has large fluxes across the 4 hrs. Does this reflect differences in incubation time, perhaps started later in the morning or afternoon than others..? Perhaps add a sentence like “all daytime incubations were started at X am, and all nighttime incubations were started at X pm”.

Ln 116: Species name must be italicised

Ln 134: Were filters changed between samples and incubation time points? If so, how? If not, could sediment and particles trapped in the filters from T0 have impacted samples at T2 and T4?

Ln 203-207: Be clear if benthic cover (dominance) was used as a fixed categorical factor in analyses. If it was, how may the difference in cover for algal dominance (72-83%) or cyanobacteria dominance (83-91%) have influenced results from these respective tents? I believe that a 10% difference in algal or cyanobacteria cover could be quite influential. This seems like an important consideration given that there were possibly just n=3 replicates per substrate type. (See major comment above).

Ln 205: Was PERMANOVA conducted on raw or model data? This is important to state. I am not sure how robust this analysis is to such low sample size n=2 within raw data, but also unsure whether such analyses should be conducted on modelled data. Further, what biochemical processes (variables) were analysed using PERMANOVA? Table S3 shows just one set of output data, but how does this translate to NCC, NCP, etc? What exactly was tested here?

Results
The results would benefit from a few subheadings to form structure. E.g. #1 general temperature, salinity, light in the tents – baseline conditions / tent effects, #2 incubations with differences in AT, CT, pH, N, etc. among tents and substrate, and #3 NEC and NCP among tents and substrates.

Ln 215-218: How did these leakages impact incubations? Given the exchange rates were greater in these two tents, is it possible that their leaking confounded the results from these tents? How was this accounted for?

Ln 227, 228, 231, etc: Again, I would stick to past tense in results section, e.g. “NCP **showed** a clear diurnal pattern”

Figure 5: This figure looks like a copy-paste model output. Minimum, panels should be better labelled. However, a more informative figure could be produced that summarises the model outputs for the five substrates, day and night.

Figure 6: Panels should be labelled with A, B, C and D. Also, it now seems that Table 1 is redundant given this information is provided here in Figure 6. It is much easier to view as a figure, so I suggest deleting Table 1.

On this note, it would be nice to see other seawater chemistry data (AT, CT, pH, etc) presented like this in a separate figure or table. I feel the results are short and overlook

baseline measurements of seawater chemistry among tents and substrate types. How did AT, CT, pH vary among substrates and across incubations?

Ln 242: “per” should be capitalised in PERMANOVA

Discussion
Ln 253: This is a nice contrast, but make it clear by stating the ranges of NCC you found, as well as that commonly found in the literature.
Ln 254: What does “no real gain in primary habitat” mean? Do you mean reef accretion / coral growth? If so, use this very carefully, as NCC and accretion are not always coupled. Low (or high) rates of NCC do not always correspond to low (or high) accretion rates. Did you measure primary habitat gain somehow, or is this assumed from NCC rates?
Ln 255: As above, saying “accumulation of biomass through photosynthesis is low” may not be true.
Also, use past tense throughout: “was” not “is”, “were” not “are”

Ln 260: I am not convinced by this statement. Fig 6 shows differences among substrates, which more explicit analyses may reveal. Functional homogenisation is quite a loaded term to use from n=2-3 replicates.

Ln 267, 280, 298, etc: Consider a new paper that addresses calcification and dissolution on dead and live coral surfaces:

- Romanó de Orte, M., Koweek, D. A., Cyronak, T., Takeshita, Y., Griffin, A., Wolfe, K., ... & Caldeira, K. (2021). Unexpected role of communities colonizing dead coral substrate in the calcification of coral reefs. *Limnology and Oceanography*

Ln 269: In reference to my comment above, how was “accretion rate” calculated?

Ln 274: “an” should be “and”
Ln 275: “Koweet” should be “Koweek”

Ln 290-293: This text is useful but has no clear point as currently expressed. I also see its relevance at Ln 365.

Ln 341: As above, this key message may not be correct given the low sample size and unclear PERMANOVA analysis. The narrative and analyses must be readdressed.

Ln 345: This level of 34-36% coral cover is very high for the Caribbean. Why was this done? Were corals intentionally moved into tents to create this high coral cover? If so, were they left to stabilise for several days after relocation? Were any metrics of coral condition measured before (and/or after) incubations? Is it possible that the corals were stressed and under-performing?